UNIVERSITÀ
DEGLI STUDI
FIRENZE

PhD in
Chemical Sciences

CYCLE XXXVI

COORDINATOR Prof. Anna Maria Papini

Bioconjugation of proteins with ^{19}F and nitroxide tags for in vitro and in cell characterization

Academic Discipline (SSD) CHIM/03

Doctoral Candidate
Dr. Vitali Valentina

(signature)

Supervisor
Prof. Banci Lucia

(signature)

Coordinator
Prof. Papini Anna Maria

(signature)

Table of Contents

1. Abstract

Proteins are complex biomacromolecules that are at the basis of all physiological processes, such as respiration, metabolism, immunity, and signaling. To better understand their function, it is essential to elucidate their structures, dynamics, and interactions. Structural biology techniques, such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, cryo-electron microscopy (cryo-EM), Förster resonance energy transfer (FRET) and Electron Paramagnetic Resonance (EPR) are essential for this task. However, to effectively utilize some of these techniques, bioconjugation with small external molecules - often referred to as tags or probes - at specific sites is essential for the functionality of the technique itself. In this thesis, we explored the application of various tag topologies, from ^{19}F moieties to nitroxide spin labels. The former were investigated to explore novel alternatives for post-expression covalent modification of tyrosine residues, providing valuable approaches for preparing ^{19}F NMR samples. On the other hand, we compared four different nitroxide spin probes bound to cysteine residues to assess their different behaviour while attached on two different secondary structural elements (α-helix and the β-sheet). At the same time, we employed a Cu(II) chelate complex, attached to a double histidine motif. Retrieving the distances between the Cu(II)-nitroxide orthogonal spin labels, using pulsed EPR techniques, while alternating both label types between the α-helix and the β-sheet allowed us to analyze in depth the nitroxides features. Lastly, one of the aforementioned nitroxide spin labels was also utilized for in-cell studies on both *Escherichia coli* and *Pichia pastoris* through EPR spectroscopy. We employed different combinations of deuterated protein and deuterated cells to assess the influence of deuteration on the transverse relaxation time T_2 (or phase memory time T_m). The presence of deuterium causes an increase of the phase memory time (due to its smaller magnetic moment compared to proton), which gives the advantage of a better sensitivity with shorter data collection times or the possibility to extend the data time over which the echo is measured generating the possibility of measuring longer distances between the nitroxides labels.

2. Introduction

In recent decades, the field of structural biology has evolved from a simple reductionist approach to a more integrated one, thus constituting a cornerstone for advancing our understanding of the intricate workings of living organisms. In the vast majority of cases, the direct relation between the function of a biomacromolecule and its structure, has promoted the determination and analysis of the three-dimensional architectures of proteins and nucleic acids to gain deep insights into the molecular aspects that regulate a wide array of biological phenomena. These phenomena involve the mechanisms behind protein folding and functionality, the processes of DNA replication and transcription, and the extensive network of interactions in the cell. Alterations in structural properties are the molecular basis of severe diseases, such as cancer and neurodegenerative conditions. To untangle valuable insights on these aspects, various structural techniques, including X-ray crystallography, cryo-electron microscopy (cryo-EM), Förster resonance energy transfer (FRET), Nuclear Magnetic Resonance (NMR), and Electron Paramagnetic Resonance (EPR), have been harnessed both *in vitro* and in-cell. Since the resolution of the 3D structure of myoglobin in the late 1950s[1], X-ray crystallography has revolutionized the field of structural biology allowing the most detailed investigations on single proteins and also on biomolecular complexes such as the ribosome. Over time, it has become an indispensable high-resolution technique and nowadays around 90% of the structures deposited in the Protein Data Bank (PDB) derive from X-ray data. Nevertheless, its bottleneck lies in obtaining high quality protein crystals. This step can be expensive and time consuming. These difficulties can partially be overcome by the employment of the more recently introduced cryo-electron microscopy, increasingly used to study biomolecular assemblies that are non-permissive to crystallisation (it requires frozen solution samples), even though the optimization of the samples requires significant efforts. FRET is appealing for elucidating crucial information about protein dynamics or interactions by measuring the distance between two sufficiently close in space fluorescent molecules previously introduced on the system under investigation. On the other hand, NMR can retrieve both dynamic and structural information on biomolecules directly in liquid solution, nevertheless, high resolution liquid state NMR spectroscopy is still generally restricted to biomolecules below 100 kDa of molecular weight. EPR is a

methodology exclusively sensitive to paramagnetic centres, *i.e.* with unpaired electron spins. It is often complementary to other structural techniques and is particularly useful for studying the structure and dynamics of biological molecules in solution and inside cells.

In this PhD research work both NMR and EPR spectroscopy techniques were exploited to interrogate aspects about protein behaviours both *in vitro* and in-cell.

2.1. Tags in structural biology

In the field of modern biology, the usage of selectively and strategically incorporated small tags to specific sites within biological molecules like proteins, nucleic acids or lipids[2,3], is very common. This approach is exploited by several techniques to retrieve valuable insights on biomolecules structure, function, dynamics and interactions, both *in vitro* and in their native intracellular environment. The proteins themselves can be modified with a variety of functional molecules, including luminescent dyes, paramagnetic probes, photo-responsive molecules or polymers. For example, among the multitude of paramagnetic tags available, nitroxide spin labels[4], with their unpaired electrons and well-defined spin states, have become indispensable in EPR. They enable the precise measurement of distances between paramagnetic tags, conformational changes, and dynamics within biomolecules. On the other hand, fluorescent tags, such as the CyDye[5], AlexaFluor[6] and ATTO series[7], have been extensively employed both in fluorescence microscopy and FRET. While the former is widely used for selective real-time detection of the localization, trafficking, and activities of proteins or nucleic acids inside cellular environment, the latter allows to acquire knowledge on molecular interactions, structural and conformational changes, and protein dynamics. In general, it is imperative that the labels attach selectively and with great accuracy only the desired amino acid residue. One of the most common approaches involves the direct chemical modification of proteins to produce novel bioconjugates while preserving the integrity and functionality of the protein. Fulfilling particular prerequisites in these chemical reactions is essential for optimal results. First, they should be carried out in water or in conditions as close as possible to the physiological ones: this includes a pH close to 7 and temperatures that can range from room temperature to a maximum of 40°C. Moreover,

the external chemical reagent added to the proteins should be - as much as possible - non-toxic, highly reactive and used in low concentration to avoid the risk of precipitation.

2.2. Amino Acid Residues as Anchors for External Tags

The choice of the residue on which the external tag needs to be attached is almost as significant as the choice of the tag itself. The nature of the amino acid functional groups is crucial for developing selective chemical reactions that can target specific amino acid residues. However, independently by its intrinsic chemical nature, the amino acid selected for the labelling should be placed on the surface, readily accessible by the solvent, rather than buried deep inside the protein core. Furthermore, labelling the selected residue should ensure minimal interference with the protein's native function without disrupting the structural integrity, while preventing the protein misfolding. Another vital feature to meticulously control is the number of potential labelling sites. In some cases, attaching the labelling agent on multiple sites can lead to convoluted and difficult to interpret data. Therefore, limiting the number of labelling sites ensures that the tag interacts with a specific amino acid, providing clear and accurate insights on the protein characteristics. Moreover, modifying the protein on several sites could increase the risk of interfering with its natural function. Controlling the numbers and the positioning of the residues of interest can be achieved through site direct mutagenesis. This is a powerful molecular biology tool that introduces specific amino acid residues in predetermined positions, while removing other undesired residues, by modifying the DNA sequence that encodes for that specific protein. It is crucial that this process is carried out considering that the protein stability and/or activity should be maintained unaltered.

In this thesis the most common residues exploited for bioconjugation of small molecules will be briefly introduced and explained. Due to the peculiar chemistry of its thiol group, cysteine has been one of the most successful residues to selectively host small molecules modification. Moreover, its low natural abundance on protein surface (below 2%)[8], makes this residue the ideal candidate for the task at hand for facile modification at a single site. Indeed, the presence of the reactive -SH group allows for precise and specific modification of proteins by undergoing disulfide exchange to form a protein-tag sulfur-sulfur bond, but also alkylation with suitable electrophiles such as α-halocarbonyls (*e.g.*,

iodoacetamide)[9] and Michael acceptors (*e.g.*, maleimides)[10]. Even though cysteine residues can be considered as the "workhorse" of chemical covalent modifications, valuable alternatives need to be analyzed. In fact, either some proteins can be characterized by essential cysteine residues directly involved in the protein native structure or function (*e.g.*, formation of disulfide bridges) and therefore cannot be labelled, or other proteins may have a number of cysteine residues too high for modifying only a low number of sites. Meanwhile, despite its high natural abundance, lysine remains a popular choice for modification due to the high number of existing methods for the modification of primary amines (even if it is worth mentioning that sometimes the N-terminus region can also be affected)[11]. Preferential conjugation to amines, rather than the nucleophilic thiol of cysteine, can potentially be accomplished by employing more reactive electrophiles such as activated esters[12], sulfonyl chlorides[13], or isothiocyanates[14]. A well-established alternative reaction is based on the reductive alkylation using aldehydes in the presence of sodium cyanoborohydride[15]. However, the high natural abundance of lysine residues poses a challenge for achieving site-selective modifications and, when the aim is to label only a few specific protein sites, it is not advisable to follow this strategy. In contrast to cysteine and lysines, the other 18 amino acids have been only minimally explored as tools for selective modification. Among these two of the most interesting ones are tyrosine and histidine residues. While the former is considered an amino acid with an amphiphilic nature (having both a hydrophobic, the aromatic ring, and a hydrophilic, the -OH group, part), the latter is a hydrophilic residues. Due to its intrinsic nature, tyrosine tends to be mostly buried deep inside the hydrophobic core of the protein, while only a small percentage of residues can be found on the protein surface, making it an ideal candidate for site selective covalent modifications. The main chemical reactions on this residue involve the use of diazonium salts[16,17], the ene type reaction with the use of diazodicarboxyamides reagents[18] and the three component Mannich type reaction[19]. All three methods will be discussed more in depth in Section 4.1.1.3 On the other hand, the histidine residue has a unique hetero aromatic imidazole side chain and play irreplaceable functional roles in peptides and proteins including metal ion coordination and hydrogen bond donor/acceptor. Current approaches for selectively modifying histidine primarily use N-substitution (alkylation or phosphorylation)[20,21] reactions involving the moderately nucleophilic imidazole group. However, this approach

is challenged by interference from lysine and cysteine residues, but enhancing selectivity for histidine labeling could be effectively achieved through the implementation of a double histidine motif (dHis)[22]. As the name suggests, this motif involves placing two histidine residues close to each other in the protein sequence and ensuring that their histidine rings align on the same side of the protein in the three-dimensional structure. Specifically, when located on an α-helix, there should be a four-residue gap between the histidines (at positions *i* and *i+4*), and when positioned on a β-sheet, a two-residue gap is appropriate (at positions *i* and *i+2*)[23]. This motif has been extensively employed for EPR applications since its combination with metal complex chelators such as CuNTA or CuIDA allows the detection of narrower distance distributions in respect to the one extracted from nitroxide-nitroxide distance (See Section 4.2.3.2.2). In conclusion, developing novel methodologies for amino acid modification is crucial to complement the already existing ones. Moreover, expanding the range of suitable residues for covalent modifications, could allow the attachment of different spin labels on the same protein (orthogonal labelling).

3. Aim of the project

Despite the vast progress in the bioconjugation field, there are still many challenges that need to be addressed, both in conjugation strategy and in the optimization of the tag selected for the conjugation. In this thesis we employed different types of small tags, from ^{19}F probes to nitroxide spin labels and copper chelators for the application in the field of magnetic resonance techniques. The focus was to characterize the behaviour of different tags after the conjugation on proteins, but also to enlarge the scenario of post-expression covalent modification approaches. The latter was pursued by developing the first labelling strategy for attaching small, fluorinated probes to tyrosine residues. We intended to address the critical gap in the current methodologies for incorporating ^{19}F nuclei into biomolecules, since the traditional residues like cysteines or lysines may not always be the optimal choice. This is crucial for widening the methodologies for preparing fluorinated proteins for ^{19}F NMR. Meanwhile, on the other hand, we aimed to evaluate the characteristics and behaviour of four different nitroxide labels through pulsed EPR spectroscopy. By retrieving the distance distributions between these nitroxides and a copper chelator agent we determined, through the width and the shape of every distribution, their performances by alternating their position on two different secondary structural elements. Lastly, we wanted to employ one of these nitroxides for in-cell EPR studies to assess the effect of a combination of protein and/or cell deuteration on the EPR spectra. We systematically studied the deuteration effects on the electron spin transverse relaxation time T_2. Deuteration successfully increased the T_2 values, which resulted to be pivotal both for increasing the sensitivity and for opening the possibility of measuring longer distances.

4. Methodological Approaches

4.1. ^{19}F NMR

^{19}F NMR stands out as an incredibly versatile and powerful technique, extremely useful in the study of biomacromolecules structure and dynamics. It has gained increasing interest [24,25] due to the high gyromagnetic ratio of the ^{19}F nucleus, its high natural isotopic abundance (almost 100%) and its nuclear spin of ½[26]. Moreover, this nucleus is almost nonexistent in biological contexts, making it an easily identifiable and unique NMR probe. Furthermore, being absent in natural biomolecules, its NMR spectra are not affected by background signals, further enhancing its potentialities in biological investigations. This allows the characterization in solution of larger biomolecules with respect to those studied through ^{1}H NMR. This technique has been widely used for detecting protein and nucleic acid conformational changes, protein-protein[27] or protein-ligand interaction[28,29], and protein folding and unfolding[30,31]. Not to be overlooked is the huge potential that this technique has shown in the field on in-cell NMR[32-34], opening the possibility to study biomolecules directly inside their intracellular environment but without the impairment of the typical severe line-broadening. This phenomenon derives from the interaction of soluble globular proteins with the cellular components[32] that, by causing an increase in the ^{1}H nuclear spin transverse relaxation, can lead in the worst cases even to the loss of all signals in ^{1}H NMR spectra.

4.1.1. Methods for Biomolecules Fluorination

Due to its essential absence in biological environment, the key element of the success of ^{19}F NMR is developing successful methods to selectively incorporate the fluorinated probe into the investigated protein. It is worth mentioning that, due to the similarity in the van der Waals radius between the hydrogen and the fluorine atoms, the ^{19}F incorporation should not disrupt the protein structure[35,36]. There are two main methods for fluorine probe incorporation. The first one is based on the direct expression of ^{19}F proteins usually by adding unnatural fluorinated amino acids into the protein expression culture medium. The second methodology involves a chemical modification on a specific protein site after

the protein expression and purification. The sections below are meant to briefly introduce both approaches and to highlight their strengths and weaknesses. (**Figure 1**)

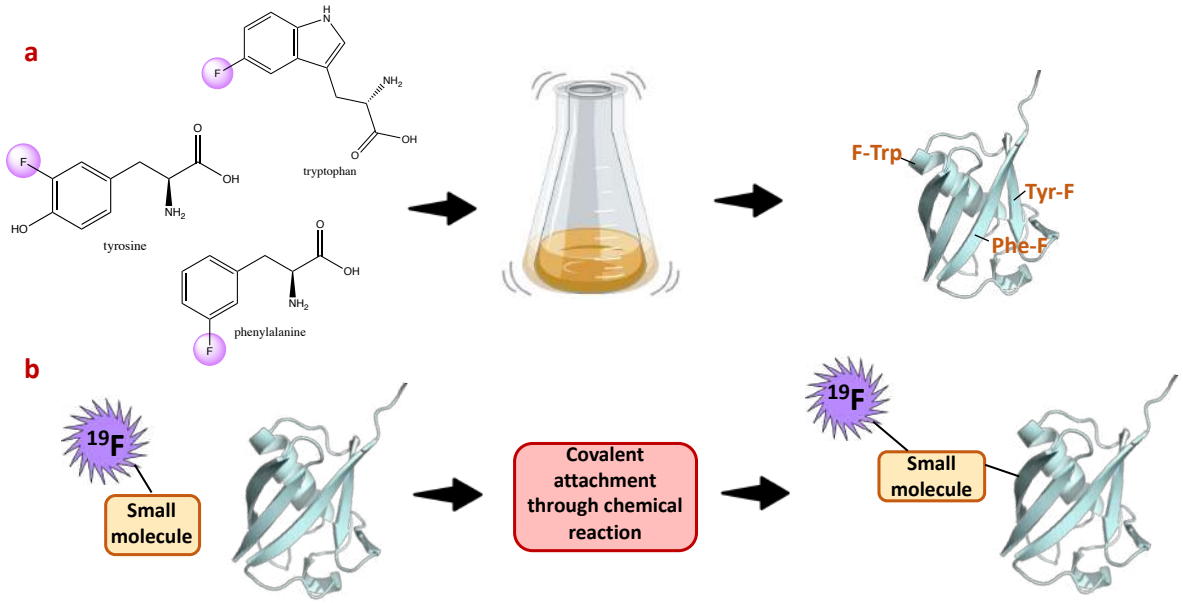


Figure 1. *Schematic representation of the two main approaches used for ^{19}F incorporation. a) Direct expression of ^{19}F proteins. Usually fluorinated analogues of amino acids (like the ones represented in this scheme) are added to the culture medium during the protein expression and the result is a protein globally substituted with ^{19}F residues. b) Post-expression approach for attaching small, fluorinated molecules through covalent bonds site specifically on the protein. Based on the selected tag and the chemical reaction of choice, the labelled residues involved are usually cysteines and lysines.*

4.1.1.1. Direct Overexpression

Direct overexpression of fluorinated proteins is generally achieved by adding fluorinated amino acids analogues in the growth medium during protein expression. The most commonly exploited mono-fluorinated amino acids are analogues of the aromatic amino acids. These include the ortho, meta, and para derivatives of phenylalanine (2-, 3-, and 4-F-Phe), the meta derivative of tyrosine (3-F-Tyr), and tryptophan fluorinated at specific indole ring positions (4-, 5-, and 6-F-Trp) (**Figure 2**).

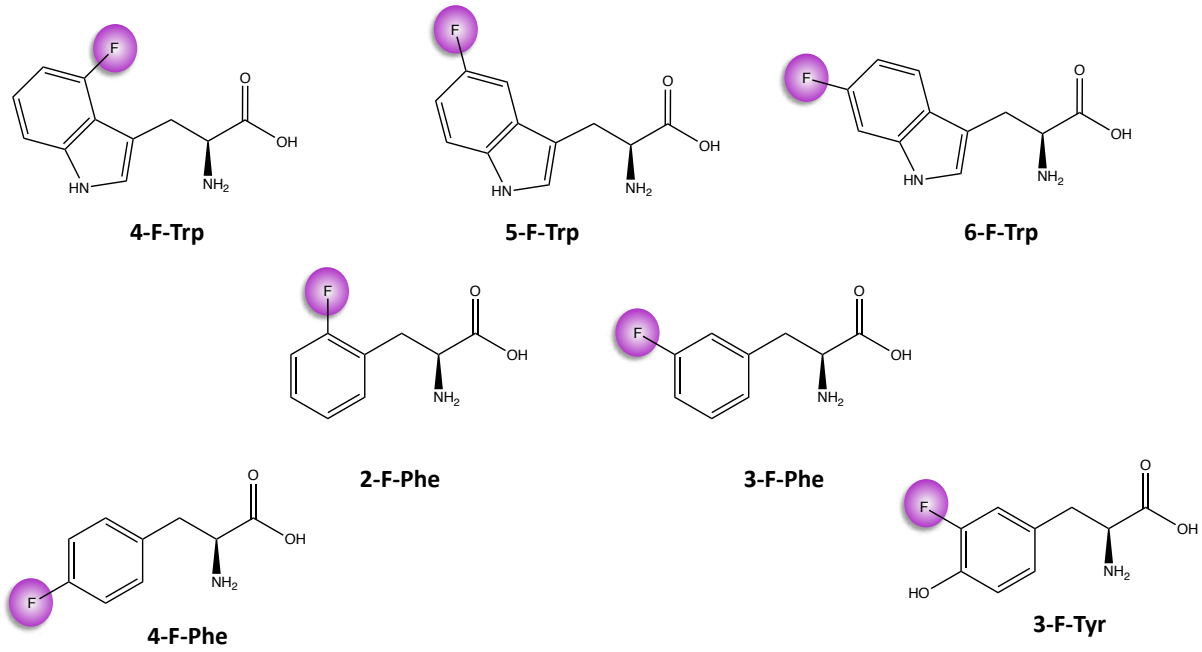


Figure 2. *Chemical structure of the most employed mono-fluorinated analogues of the aromatic amino acids: tyrosine, tryptophan and phenylalanine.*

A successful incorporation requires the addition of glyphosate to the growth medium, which inhibits the synthesis pathways of aromatic amino acids[37] to ensure that the fluorinated analogs replace the natural amino acids. In this case, the endogenous aminoacyl-tRNA synthetases are tasked with charge their paired tRNAs with ^{19}F modified amino acids. However, when this approach is employed, the fluorinated residues consistently replace the natural amino acids in all the locations where the original amino acid was present. This leads to a global labelling of the protein which could result in some side effects such as structure disruption and possible ^{19}F spectral overcrowding[38]. At the beginning of the 2000s, a new technique was introduced, with the aim of overcoming the aforementioned issue. This approach relies on the introduction of an orthogonal amber suppressor tRNA with a paired tRNA synthetase with the possibility to introduce a broader variety, not only aromatic, of ^{19}F-labelled acids at single site-specific positions[39] (**Figure 3**).

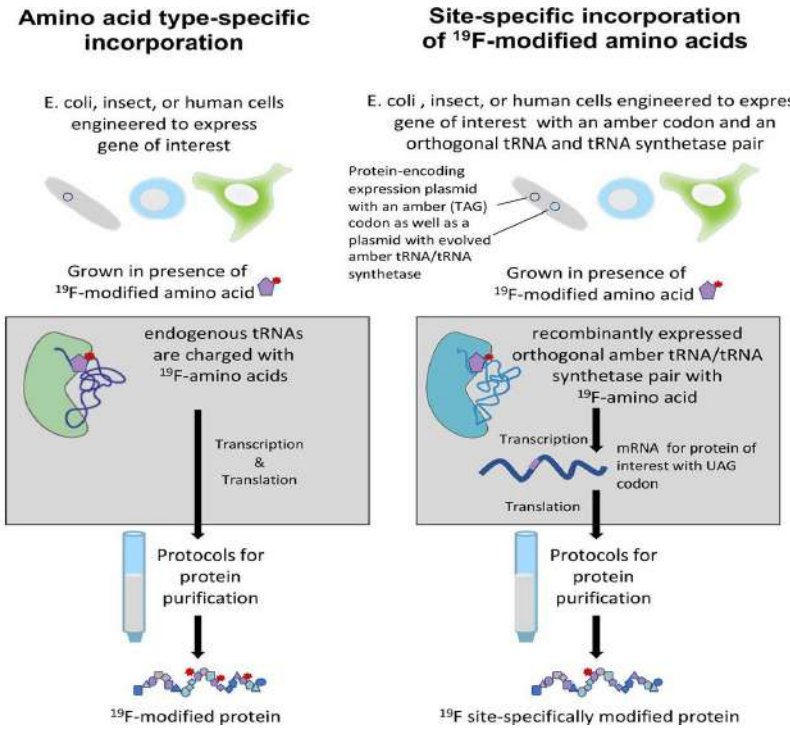


Figure 3.*On the left, biosynthetic amino acid type-specific incorporation of ^{19}F modified amino acids. On the right, site-specific incorporation of ^{19}F modified amino acids using recombinantly expressed orthogonal amber tRNA/tRNA synthetase pairs. Figure adapted from the Gronenborn review on ^{19}F*[40]

4.1.1.2. Post-expression Modification

While biosynthetic labeling is well-established for many fluorinated amino acid analogs, several challenges need to be addressed in some cases, including a reduced expression yields due to the toxicity associated with the fluorinated precursors, perturbations deriving from the incorporation of the fluorinated amino acid that can lead to protein instability or misfolding, and a reduced uptake by the aminoacyl-tRNA synthetase[41]. For these reasons, exploiting an approach that involves the formation of a covalent bond between the protein residue of choice and a fluorinated chemical reagent, after the protein expression and purification, is an attractive a potentially low-cost alternative. Meanwhile, it is also important to consider that this methodology can be more challenging in controlling the position and orientation of the fluorine-labeled probe on the protein. Nevertheless, several examples have been reported of post-expression conjugation on cysteines, lysines or tyrosines, either native or introduced with mutagenesis. To avoid the precipitation of the protein during the labelling step, the conditions need to be mild: the reaction should be carried out on aqueous solutions or in buffers as close as possible to physiological conditions. The pH should be kept around 7

and the temperature should be at maximum around 42°C. Regardless of the type of the reaction, or of the kind of amino acid residue involved, an overall precise evaluation and a thorough testing for extensive optimization are essential for each protein and every substitution involving a specific fluorinated tag. The quantification of labeling extent is typically done using mass spectrometry, like we did extensively in our research work. The sections below introduce the most common fluorinated tags that are currently exploited for cysteine and lysine post-expression conjugation. Afterwards, a brief section on tyrosine conjugation is also introduced.

4.1.1.2.1. Types of ^{19}F tags

The strategy that is usually applied for site-specifically bioconjugation of small molecules involves the employment of different chemical reactions for each amino acid residue. As already discussed, for the unique chemistry of their side chains, cysteines and lysines usually play a key role in post-expression covalent modification. There are plenty of examples where solvent accessible thiol groups of cysteines easily react with small, fluorinated molecules forming a new covalent bond. Specifically, reagents such as 3-bromo-1,1,1-trifluoroacetone (BTFA) or its analogues such as 2-bromo-N-(4-(trifluoromethyl)phenyl)acetamide (BTFMA) or N-(4-bromo-3-(trifluoromethyl) phenyl)acetamide (3-BTFMA) have been often employed[42-46]. Other cysteine-related small tags involve the 2,2,2-trifluoroethanethiol (TFET) that binds the cysteine through the formation of a disulfide bond[46-48] and the 4-(perfluoro-tert-butyl)phenyliodoacetamide (PFP)[49]. Another fluorinated cysteine tag worth mentioning is the 1-(4-(trifluoromethyl)benzyl)-1H-pyrrole-2,5-dione (TFBDP) which, by exploiting the already well established maleimide-thiol reaction chemistry, has shown high labelling efficency[46]. Lysine side chains have been successfully labelled both with n-trifluoroacetylcaproxyloxy succinimide ester (TFCS), a common crosslinker that reacts under mild reaction conditions (pH below 7.5), and fluorinated indoles and indole-derivatives[50] (**Figure 4**).

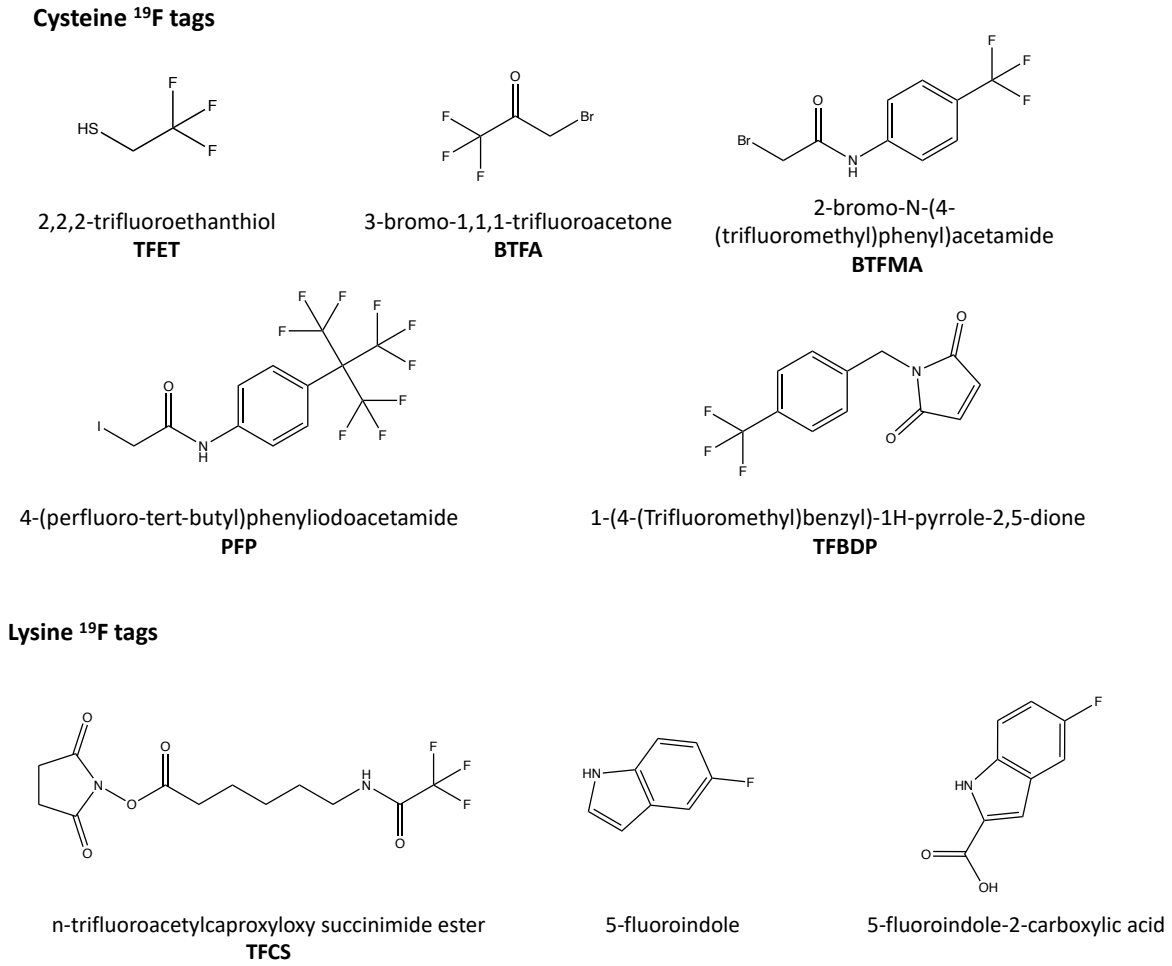


Figure 4. *Chemical structure of the most used [19]F tags both for cysteine and lysine residues covalent modification.*

As previously discussed, tyrosine represents an attractive alternative to cysteine and lysine as a potential labelling site. Its low natural abundance and its amphiphilic nature ensures that the number of tyrosine residues on the surface is quite low. Therefore, by selecting the right reaction conditions, it can be possible to covalently label the relatively rare surface exposed tyrosines in a site-selective way. To the best of our knowledge, no previous work has been conducted for investigating post-expression bioconjugation of small ^{19}F tags on tyrosine residues.

4.1.1.3. Tyrosine conjugation

In this thesis, we focused on developing a new strategy to selectively incorporate ^{19}F moieties to tyrosine trough post-expression covalent modification, adjusting the already known three-component Mannich type reaction[19]. In this section we briefly introduce the

pros and cons of the most famous approaches for conjugating small molecules or small peptides on tyrosine residues.

One fruitful approach for tyrosine conjugation is the use of diazonium reagents. They have been successfully exploited for the modification of the capsids of both tobacco mosaic virus (TMV)[17] and bacteriophage MS2[16], with efficiency above 90% with short reaction times, but using high pH values (pH = 9). Even though in these cases the reaction showed a high chemoselectivity, in other examples clear manifestations of cross reactivity, especially towards histidine, have been observed[51]. To address this aspect, a new methodology was used in which the pH is decreased to 4.5. However, these adjustments caused a slowing of the reaction rate and relatively high reduction of the bioconjugation yield only compensated with the use of 20-fold (in respect to the protein) of diazonium salts[52]. Afterwards, the development of stable diazonium salts, such as 4-formylbenzene diazonium hexafluorophosphate, removed the necessity to produce *in situ* the diazonium reagents, allowing a better control over the equivalents used, reducing the potential for side reactions[53]. Another vital approach for tyrosine bioconjugation involves the use of cyclic diazodicarboxyamides, such as 4-phenyl-3H-1,2,4-triazoline-3,5(4H)-diones (PTADs). These compounds react faster than the diazonium reagents, with reaction times typically between 15 to 30 min, and are selective for tyrosine residues[54,55]. In general, the reaction can be carried out over a wide pH range (2–10), although higher yields are generally obtained at higher pH values. However, unluckily, in water PTAD reagents can decompose to isocyanate, which in turn can cause side reactions at lysine residues or at the N-terminal of the protein. This issue was easily circumvented by using Tris buffer, which acts as a scavenger for isocyanate[56]. As mentioned before, the reaction used for our investigation is the three-component Mannich type reaction. This approach targets the phenolic side chain of tyrosine residues. The reaction involves a first step where an *in situ* imine condensation between an aldehyde and an electron-rich aromatic amine takes place. Afterwards, the phenol ring of the tyrosine is deprotonated and undergoes an electrophilic aromatic substitution with the iminium ion, resulting in the formation of a novel carbon-carbon bond (**Figure 5**). Theoretically, with this kind of reaction it is possible to add two different functionalities at the same time on the tyrosine ring through the aldehyde functional group. However, it was demonstrated that formaldehyde is the most reactive among all the aldehydes, while crotonaldehyde and 2-furaldehyde have a

decreased reactivity in this reaction and others such as benzaldehyde and propionaldehyde are not reactive at all[19].

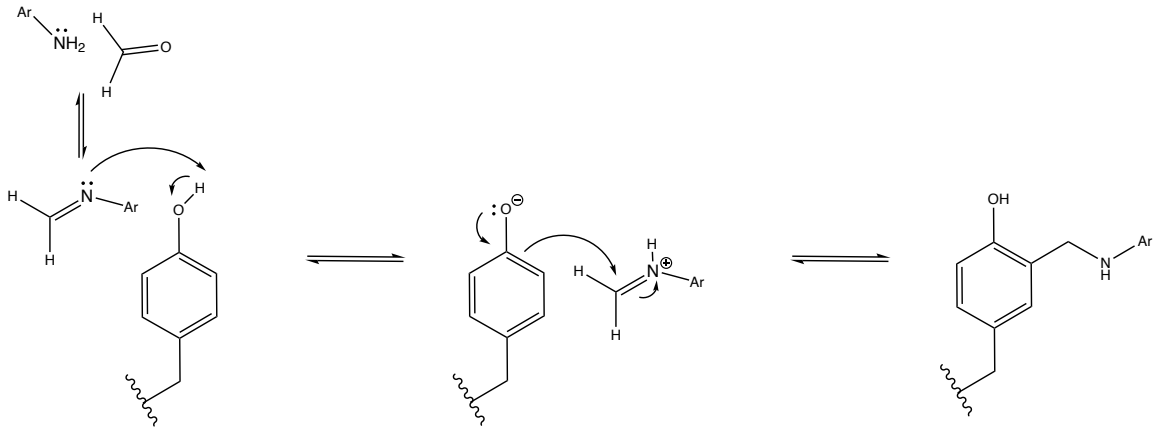


Figure 5. *General scheme of reaction of the three-component Mannich type reaction. The first step involves the reaction between an aldehyde and an aromatic amine (Ar) with consequent formation of an immine. The phenol ring of the tyrosine is deprotonated and an electrophilic aromatic substitution between the tyrosine and the iminium ion takes place. The result is the formation of a novel carbon-carbon bond in orto position on the tyrosine aromatic ring.*

Instead, the aromatic amine reacts with higher efficiency if it bears electron-donating groups such as alkyl or methoxy, since aromatic amines with strongly electron withdrawing substituents, such as nitro or carboxylic, have been shown to produce little or no bioconjugation. This reaction has been successfully used both for grafting fluorophores[19] and synthetic peptides[57] to chymotrypsinogen. Recently, it has also been used for successful modification of tyrosine residues with nitroxide radical 4-amino-2,2,5,5-tetramethyl-3-imidazo-line-1-yloxy (NOX), in the field of Site Direct Spin Labelling (SDSL) for EPR purposes[58].

4.2. EPR contributions to structural biology

A brief historical footing of Electron Paramagnetic Resonance (EPR) applications in structural biology is provided in this section to extensively report the intertwined role of modern EPR in structural biology and investigating structure-function relationship.

The first contributions of EPR to biomolecular studies may be tracked back to the mid 1950s where continuous wave EPR (CW-EPR) was employed to gain information on the local environment of free radicals within proteins generated by X-irradiation of different amino acids[59,60]. During the same years, CW-EPR was successfully employed to study metal center proteins, such as human transferrin (containing Fe^{III}) and avian conalbumin (containing Cu(II)), with the aim of obtaining insights into the mode of binding of the metal ions and the type of interaction between the active sites[61]. The introduction of the first stable methanethiosulfonate nitroxide radical (1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate, shortly called MTSL) for labelling[62], and, concomitantly, the emergence of complex pulse sequences focused on the detection of dipolar couplings[63] to facilitate the distance measurements between 2-10 nm, greatly contributed to the development of EPR as a tool for structural investigation. Furthermore, the introduction of the thiol-based Site Direct Spin Labelling approach pioneered during the late 1980s by Hubbell and co-workers[64-66] allowed the introduction of small spin-bearing molecules at specific and strategically placed sites within the native diamagnetic system, enhancing the scope of EPR for structural determination[67]. At this point further development of pulse sequences[68-71] and of data processing methods, such as Tikhonov regularization[72,73], were vital to obtain detailed insights on distance distribution and consequently provide information on conformational sub-ensembles in frozen solution[74]. Both Continuous Wave (CW) and Pulsed EPR have demonstrated to be established and valuable techniques in the study of protein structures and dynamics. Indeed, CW-EPR is extremely useful in analyzing backbone and conformational dynamics[75-77] due to the sensitivity of the spectral lineshape to the local environment of the unpaired electron. On the other hand, pulsed EPR has been proven especially fruitful in providing valuable insight into structural and dynamic conformation of biomacromolecules such as proteins[78-81] and nucleic acids[82-84], contributing to conformational studies[85,86] and to gain important insights into mechanisms of complex

biomolecular machinery[87,88]. Recently, high interest has sparkled for investigating the structure and the dynamics of biomacromolecules directly in their intracellular environment, and EPR has once again showed astonishing potential in fulfilling this objective. In fact, the highly intricate intracellular environment seems to strongly affect the distance distribution when comparing the *in vitro* and in-cell samples[85,89,90].

4.2.1. The Spin Hamiltonian

The energy terms involving electron and nuclear spins can be described by the spin Hamiltonian, $\hat{H}_0$, which has the following terms:

$$\hat{H}_0 = \hat{H}_{EZ} + \hat{H}_{HF} + \hat{H}_{NZ} + \hat{H}_{ZFS} + \hat{H}_{EX} + \hat{H}_{DD} + \hat{H}_{NQ} \quad (1)$$

where H_{EZ} is the electron Zeeman interaction term, that reflects the interaction between the electron and the external magnetic field. Similarly, H_{NZ} is the nuclear Zeeman interaction term, which describes the interaction of the nuclear spin with the applied magnetic field. H_{ZFS} is the zero-field splitting term arising from interactions between different unpaired electron spins ($S > ½$) in the same atom. H_{HF} is the hyperfine coupling term which describes the interaction between an electron spin and a nuclear spin I. H_{NQ} is the nuclear quadrupolar interaction term, describing the multiple nuclear spins present in the same nucleus, *i.e.* it occurs for nuclear spins $I \geq 1$. H_{EX} is the Heisenberg exchange coupling term and describes a through-bond interaction between unpaired electron spins. H_{DD} is the dipole-dipole interaction term and describes a through-space interaction between unpaired electron spins.

A system containing a single unpaired electron, ($S = ½$), coupled to a nucleus, with a single nuclear spin ($I = ½$), is described by the electron Zeeman term, by the nuclear one and by the hyperfine interaction.

4.2.2. Continuous Wave EPR

CW-EPR is based on irradiating the sample with a low-power microwave source at fixed frequency, while the external magnetic field B_0 is swept until the resonance condition is fulfilled. Different frequencies such as X-band (9.5 GHz), Q-band (34 GHz) and W-band

(94 GHz) can be employed for recording EPR spectra. Generally, the microwave (MW) bridge is the place where MW radiation is generated and directed towards the sample, typically placed within a resonator or cavity, using a waveguide. Ideally, all microwave power should be absorbed in the resonator. This condition is referred as "critically coupled" and is an essential prerequisite for a successful cw-EPR experiment. However, the introduction of the sample causes a detuning of the resonator with a consequent reflection of microwave power that is transmitted to the detector (MW diode). Recording the reflection power against the magnetic field generates continuous wave EPR spectra. The signal-recording channel is a central component of any spectrometer and typically includes a lock-in detector and modulation amplifier, both crucial for an accurate and sensitive detection. Briefly, a modulation coil generates a magnetic field oscillation, typically at a frequency of around 100 kHz, parallel to the principal magnetic field (B_0), with a certain amplitude (B_{mod}) with the aim of modulating the EPR signal. At this point the lock-in detector recognizes and processes the modulated signal with this reference frequency, producing an output that corresponds only to the EPR signal isolated from the noise. The consequence of the field modulation is that the slope of the EPR spectra, which represents the first derivate of the absorption spectrum, is measured. In general, for samples in solution, the molecule tumbles really fast and the EPR spectra show effects of the orientation averaging the g-value and the hyperfine coupling. On the other hand, for frozen solutions or powder samples, all the orientations of the paramagnetic center with respect to the magnetic field are equally probable. In this case anisotropy is not averaged by molecular motion and the detected signal is represented by the sum, over all the possible orientations, of the g and A values. For single crystals samples, the molecule shows high dependence on the orientation of the molecule with respect to the magnetic field B_0, due to anisotropy of both the g-value and the hyperfine coupling.

CW-EPR spectra of nitroxide radicals have been widely used to monitor the difference in the local dynamics of free nitroxide spin labels and the same labels successfully attached to the desired cysteine residues after the Site Direct Spin Labelling reaction (Section 4.2.4). Nitroxide radicals are considered among the most exploited paramagnetic species for EPR. These molecules (**Figure 6**) are characterized by the presence of a nitrogen-oxygen bond (N-O), with the unpaired electron delocalized between these atoms. Since the ^{14}N nucleus possess a nuclear spin of I = 1, the hyperfine interaction between the

nitrogen nucleus and the unpaired electron, which gives rise to three different transitions, needs also to be taken into account. The EPR spectra derived from these moieties in different conditions can be easily analyzed with the help of simulated spectra[91], that can assist in better understanding on the role of the g and A tensors (**Figure 6**).

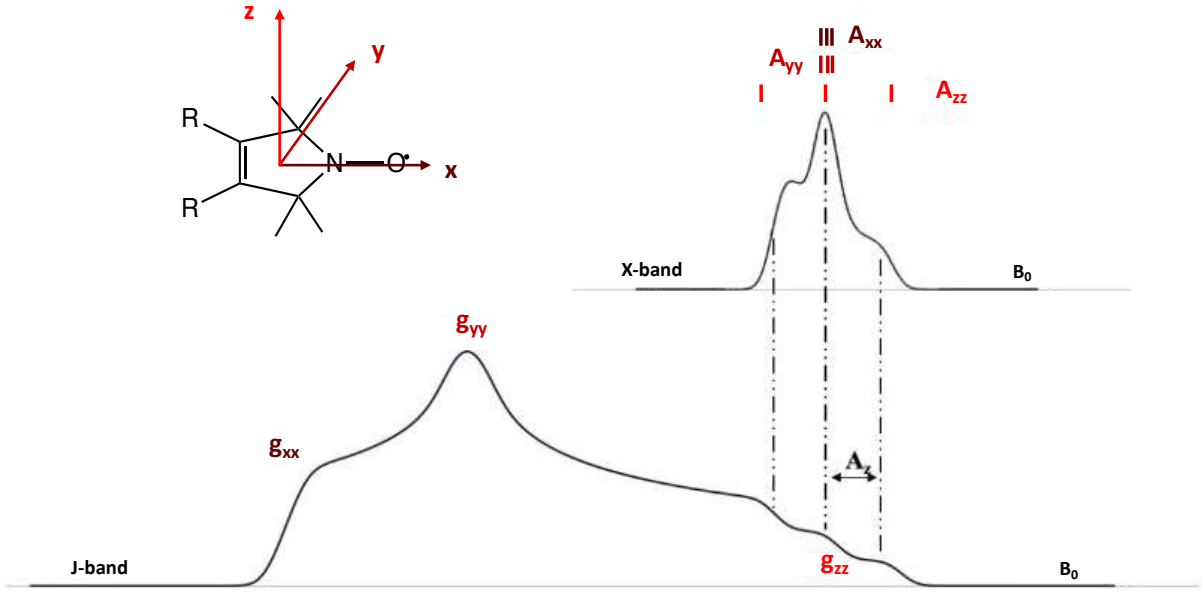


Figure 6. *On the top left, the general chemical structure of a nitroxide radical within the Cartesian coordinate system. For the nitroxides the g-tensor and the A-tensor are collinear. The rest of the image shows two simulated echo detected spectra for a nitroxide radical at two different bands: X-band (9 GHz) and the high field J-band (275 GHz). The X-band spectrum is dominated by the large hyperfine splitting of A_{zz}, while A_{xx} and A_{yy} are hidden behind the inhomogeneous broadening. The spectrum of the same nitroxide but at high fields is dominated by the anisotropy of the g-factor, with the three component g_{xx}, g_{yy} and g_{zz} well resolved.*

For example, in solution, the nitroxide EPR spectrum is affected by the thumbling and the rotational correlation time τ_c (the average time takes a molecule to tumble an angle of one radian). To illustrate the effect of nitroxide thumbing to the EPR spectrum, it is possible to simulate the spectra in solvents with different viscosity (**Figure 7**).

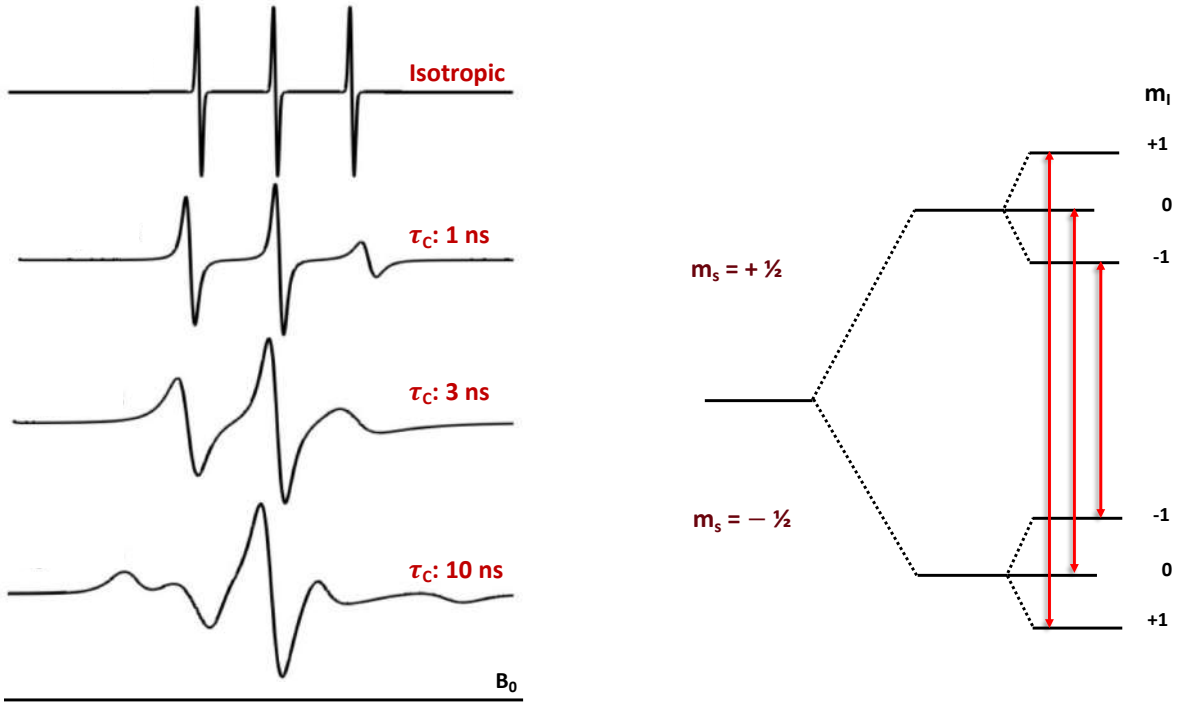


Figure 7. *On the left, EasySpin simulated cw-EPR spectra of a nitroxide radical with different correlation times (τ_c). The isotropic behaviour represents the free nitroxide label in solution, while increasing the correlation time leads to spectral broadening. On the right, a schematic representation of the energy levels of a system with a single unpaired electron ($S = ½$) and a nucleus with a nuclear spin $I = 1$ (like for the nitroxide radical). The first splitting between the two energy levels derives from the interaction between the unpaired electron and the magnetic field. The second splitting concerns the hyperfine interaction between the unpaired electron and the ^{14}N nucleus. In red are highlighted the allowed EPR transitions, according to the rule: $\Delta m_s = \pm 1$ and $m_I = 0$[4].*

In the isotropic scenario, the EPR spectrum displays three equally intense and evenly spaced peaks. This pattern is typical of free radicals, which exhibit a very short correlation time, indicative of rapid tumbling motions. When the viscosity of the solvent is increased, or the motion slows because the nitroxide is no longer free but conjugated to biomolecules, anisotropic interactions are no longer averaged out, resulting in broader and asymmetric spectra that begin to resemble the rigid-limit spectrum observed in powdered samples. Therefore, employing cw-EPR with nitroxide-based labels at room temperature enables a detailed understanding of local dynamics without being constrained by molecular size limitations. In this thesis cw-EPR has been extensively used for investigating the local dynamics of nitroxide spin labels both *in vitro* and in-cell.

4.2.3. Pulsed EPR Spectroscopy

This section is intended as a brief overview of semi-classical EPR theory, particularly of electron-electron dipolar coupling and the PELDOR and RIDME experiments, to

introduce the unique role of pulsed EPR spectroscopy in the structural biology filed. These approaches allow to study the width of conformational ensembles for the spatial organization of biomolecular systems, by measuring the distance distribution between at least two paramagnetic labels attached to the investigated biomolecule. The distance distributions are extracted from the dipole-dipole coupling between the labels, where the primary data are acquired into the time domain and then converted into a distance distribution. The EPR theory herein described can be found in several textbooks[92-94] and review articles[95-98].

4.2.3.1. The Dipole-Dipole Interaction

For systems with more than one paramagnetic center, the effect of the coupling between two interacting electron spins (S_A and S_B) is described by the H_{EX} and the H_{DD} terms already introduced in equation (1). For an isolated spin pairs (S = ½):

$$\hat{H}_0(\hat{S}_A,\hat{S}_B) = \hat{H}_0(\hat{S}_A) + \hat{H}_0(\hat{S}_B) + \hat{S}_A\mathbf{J}\hat{S}_B + \hat{S}_A\mathbf{D}\hat{S}_B \qquad (2)$$

where: $H_o(\hat{S}_A)$ and $H_o(\hat{S}_B)$ are the individual spin Hamiltonians, while ***J*** and ***D*** correspond to the exchange coupling and dipole-dipole tensors. From now on, the exchange couplings will not be considered since if the two paramagnetic centers are not linked by a continuous chain of conjugated bonds, exchange coupling is rarely resolved at distances larger than 1.5 nm. In any case, at such long distances exchange coupling is much smaller than the dipole-dipole coupling between the two unpaired electrons if the system is not conjugated.

Pulsed EPR experiments are mostly devoted to detect the effects of the dipolar coupling between two paramagnetic labels for the obtainment of the distance distribution between them (**Figure 8**).

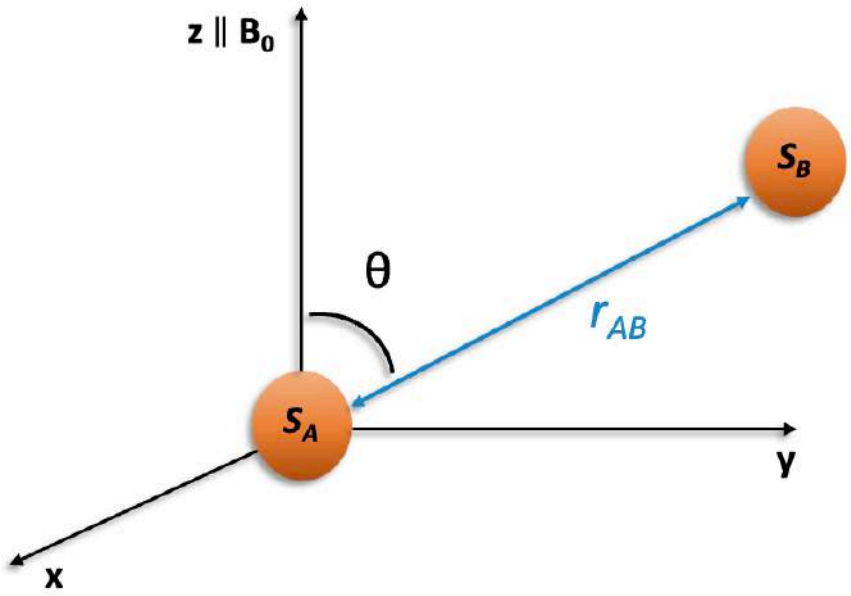


Figure 8. *Schematic representation of two spins, A and B, coupled through dipole-dipole interaction. The interspin vector that connects the two spin is labelled as r_{AB}. The angle θ describes the relative orientation of the interspin vector with respect to the external magnetic field B_0 that in this case is aligned along the x axis of the laboratory frame.*

The energy of this interaction can be described according to classical physics as:

$$E = \frac{\mu_0}{4\pi}\left(\frac{\vec{\mu_A}\cdot\vec{\mu_B}}{r^3_{AB}} - \frac{3(\vec{\mu_A}\cdot\vec{r_{AB}})(\vec{\mu_B}\cdot\vec{r_{AB}})}{r^5_{AB}}\right) \quad (3)$$

On the other hand, in quantum mechanical terms, the dipole-dipole interaction term between electrons A and B can then be defined as:

$$\hat{H}_{AB} = \frac{1}{r^3}\frac{\mu_0}{4\pi\hbar} g_A g_B \beta_e{}^2 \left[\hat{S}_A\hat{S}_B - \frac{3}{r^2}(\hat{S}_A r)(\hat{S}_B r)\right] \quad (4)$$

where r is the vector that connects spins A and B, μ_0 is the vacuum permeability constant, β_e is the Bohr magneton, and g_A and g_B are the corresponding g-tensors of the spins A and B. If the orientation of the interspin vector r is expressed in spherical coordinates, the above equation can be converted into:

$$\hat{H}_{AB} = \frac{1}{r^3}\frac{\mu_0}{4\pi\hbar} g_A g_B \beta_e{}^2 [A + B + C + D + E + F + G] \quad (5)$$

where A, B, C, D, E and F can be expresses as:

$$A = S_z^A S_z^B \,(1 - 3\cos^2\theta)$$

$$B = -\frac{1}{4}(S_+^A S_z^B + S_z^A S_+^B)(1 - 3\cos^2\theta)$$

$$C = -\frac{3}{2}(S_+^A S_z^B + S_z^A S_+^B)\sin\theta \cdot \cos\theta \cdot e^{i}\varphi$$

$$D = -\frac{3}{2}(S_-^A S_z^B + S_z^A S_-^B)\sin\theta \cdot \cos\theta \cdot e^{i}\varphi \qquad (6)$$

$$E = -\frac{3}{4}S_+^A S_+^B \sin^2\theta \cdot e^{-2i\varphi}$$

$$F = -\frac{3}{4}S_-^A S_-^B \sin^2\theta \cdot e^{2i\varphi}$$

where S_z, S_-, S_+ are spin operators, and θ is the angle that describes the orientation of the interspin vector with respect to the external magnetic field B_o. Under the high field-approximation only the secular term (A) and the pseudo-secular term (B) contribute to the dipole-dipole interaction. If the weak-coupling approximation is met as well, the dipolar term is described only by the secular term A. The weak coupling approximation can be applied if the dipole-dipole interaction (ω_{DD}) between the electrons A and B is small compared to the resonance difference between them ($\Delta\omega_{AB} = \omega_A - \omega_B$) (**Figure 9**).

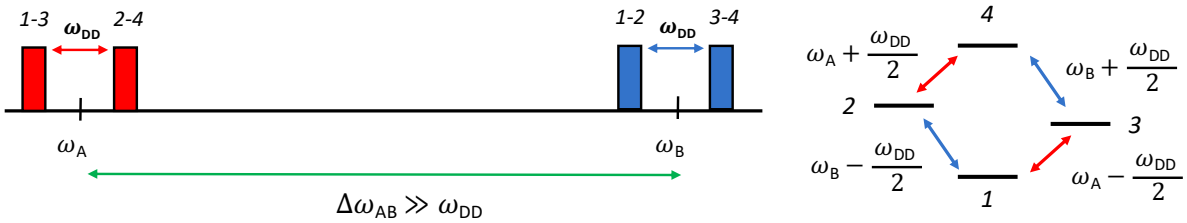


***Figure 9**. On the left, a schematic representation of the weak coupling approximation, with the dipolar splitting of the A and B spins around their respective Larmor frequencies, ωA and ωB. On the right, the corresponding energy level diagram, with the corresponding transition energies indicated.*

The frequency of the electron-electron dipolar interaction is then described as:

$$\omega_{DD} \approx \frac{\mu_0 \beta_e^{\,2} g A g B}{4\pi\hbar}\frac{1}{r^3}(3\cos^2\theta - 1) \qquad (7)$$

As stated above, in a powder sample or in the case of a frozen solution, all random molecular orientations are present and not averaged. In these cases, the angular dependency of dipolar frequency generates a characteristic spectrum called Pake pattern (**Figure 10**), which represents the resulting distribution of the dipolar frequency in the powder sample. This pattern arises because the corresponding intensities of the orientations are weighted by $\sin\theta$, and so orientations perpendicular ($\theta_\perp$) to B_0 dominate the Pake pattern, with a splitting corresponding to $\nu_\perp$. Singularities are also present at orientations parallel ($\theta\parallel$) to B_0, with a splitting corresponding to $\nu\parallel$, which is often referred to as the 'double frequency component', since $\nu\parallel = 2\nu_\perp$.

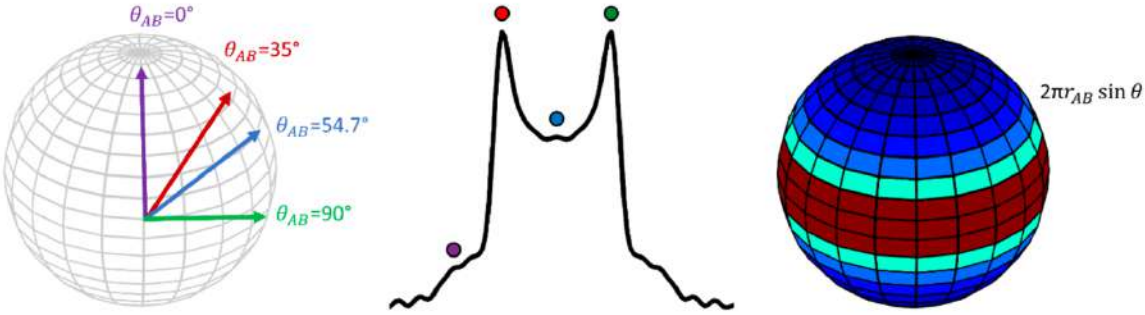


Figure 10. *On the left, a unit sphere indicating the angle θ between the magnetic field and the distance vector (r_{AB}). In the middle, a typical Pake pattern frequency response, indicating the singularities corresponding to different values of θ (shown as dots). On the right, the unit sphere with a heat-map to indicate the relative intensities of given conformations, weighted with $\sin\theta$. There is greater intensity at the equator than at the poles of the sphere, reflected by the frequency response.*

4.2.3.2. DEER and RIDME experiments for distance measurements

The initial approach employed for assessing dipolar interactions between two paramagnetic centers at close distance (below 2 nm) involves the use of continuous-wave EPR, which can provide information about the dipolar splitting. The dipolar coupling can be straightforwardly measured through cw-EPR if it is sufficiently large to split an EPR line beyond its intrinsic linewidth. However, if the dipolar splitting is smaller than the intrinsic linewidth, deconvolution methods are necessary to separate the dipolar contribution from other effects. Although deconvolution techniques are generally reliable, they can encounter stability issues when applied to spectra with unresolved dipolar peaks, requiring robust signal-to-noise ratios to avoid artifacts. Meeting these

conditions can sometimes be challenging. In contrast, pulsed EPR methods enable the direct measurement of dipolar coupling extent and, consequently, the distances between two paramagnetic centers at distances typically between 2 and 10 nm (but in some cases even beyond[99]), using specific pulse sequences. While as mentioned above cw-EPR involves continuous microwave irradiation at a fixed frequency, pulsed EPR experiments employ short, high-power microwave pulses in the nanosecond range to excite spins and perturb electron spin transitions. One of the simplest pulse sequences is the two-pulse Hahn Echo sequence.

At the beginning of every pulse sequence the macroscopic magnetization is always aligned parallel to the z axis. The application of a microwave pulse along the x axis creates an additional magnetic field B_1 perpendicular to the static field B_0, which causes spin transitions and therefore changes the resulting magnetization. In the case of the 2-pulse Hahn Echo sequence (**Figure 11**) the macroscopic magnetization is initially aligned along the z-axis (point a). Following the application of a π/2-pulse along the x-axis the magnetization is tilted and results aligned along the -y-axis (point b). During the first time delay τ_1 the spins evolve on the xy plane with different resonance frequencies due to local different magnetic fields, resulting in their dephasing (point c). Afterwards, a π-pulse inverts the spin packets (point d) and after another time delay τ_1, of the same length as the first one, the Hahn Echo (HE) is formed (point e).

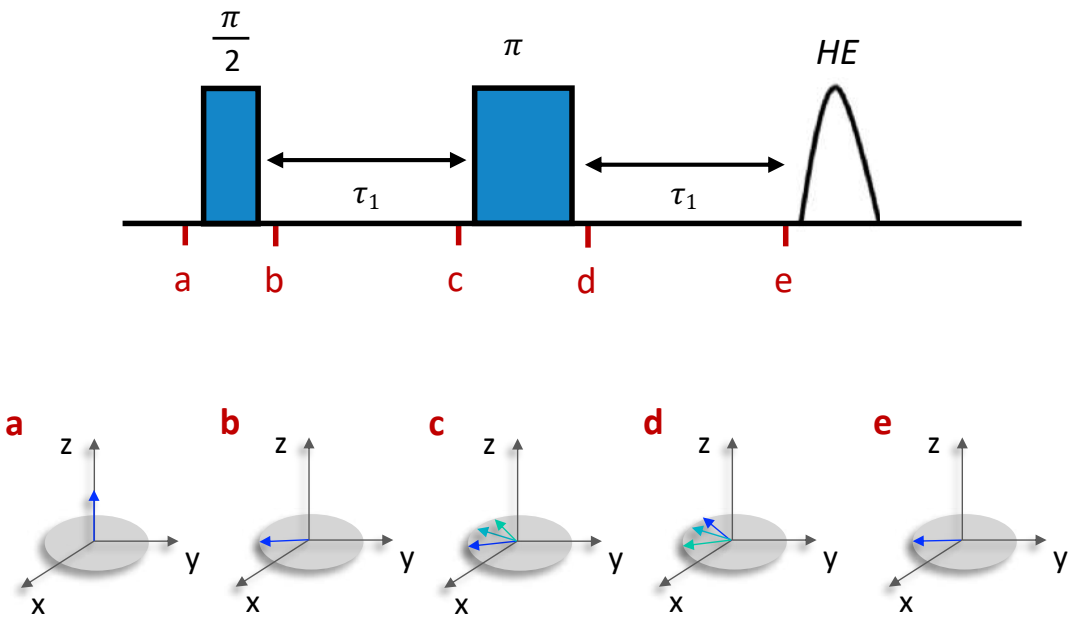


Figure 11. *On the top, the 2-pulse Hahn Echo pulse sequence. It is often used as the observer subsequence in many other pulse EPR experiments. On the bottom, a schematic representation of the spin packets behaviour through the whole pulse sequence.*

Some EPR pulse sequences have been specifically developed for the detection of dipolar couplings: the four-pulse PELDOR (or DEER) and five-pulse RIDME experiments. The four-pulse PELDOR experiment[70,71] (**Figure 12 a**) is a dead-time free version of the three-pulse PELDOR experiment[100] (**Figure 12 b**), wherein the initial π/2-pulse is replaced by the pulse train (π/2-τ_1-π-τ_1), which forms a Hahn echo at time $2\tau_1$. Briefly, the three-pulse experiment is composed by the Hahn Echo sequence with a long evolution time τ_1 between the two pulses and a π pulse at the second frequency (ν_B) which induces the transition of the spins of a different paramagnetic center characterized by a different resonance frequency. The aim is to probe the dipolar interaction between the two-spin systems A and B. However, this sequence suffered from the presence of the dead-time due to the incapability of applying the π pulse at ν_B frequency at the zero time (when the π/2 pulse at ν_A frequency is applied). This was overcome by the introduction of the four-pulse PELDOR experiment, where an additional refocusing echo is introduced. In this case the π-pulse at frequency ν_B can be applied at $t \leq 2\tau_1$, so that the pulse position is incremented through the maximum of the PELDOR signal, to accurately determine the zero-time. A final π-pulse is applied at the detection frequency at time τ_2, generating the refocused echo at time $2\tau_2$. As already introduced, the π-pulse at the second frequency is on-resonance for the spin packet B and inverts its magnetization, transferring the perturbation, through the dipolar coupling, to the other spin packet A (which is on-resonance with the detection pulses at the first frequency). This process leads to a shift in the Larmor frequency of the A spins, which no longer refocus at time $2\tau_2$, as their precession frequency is shifted by $\pm\omega_{dd}$.

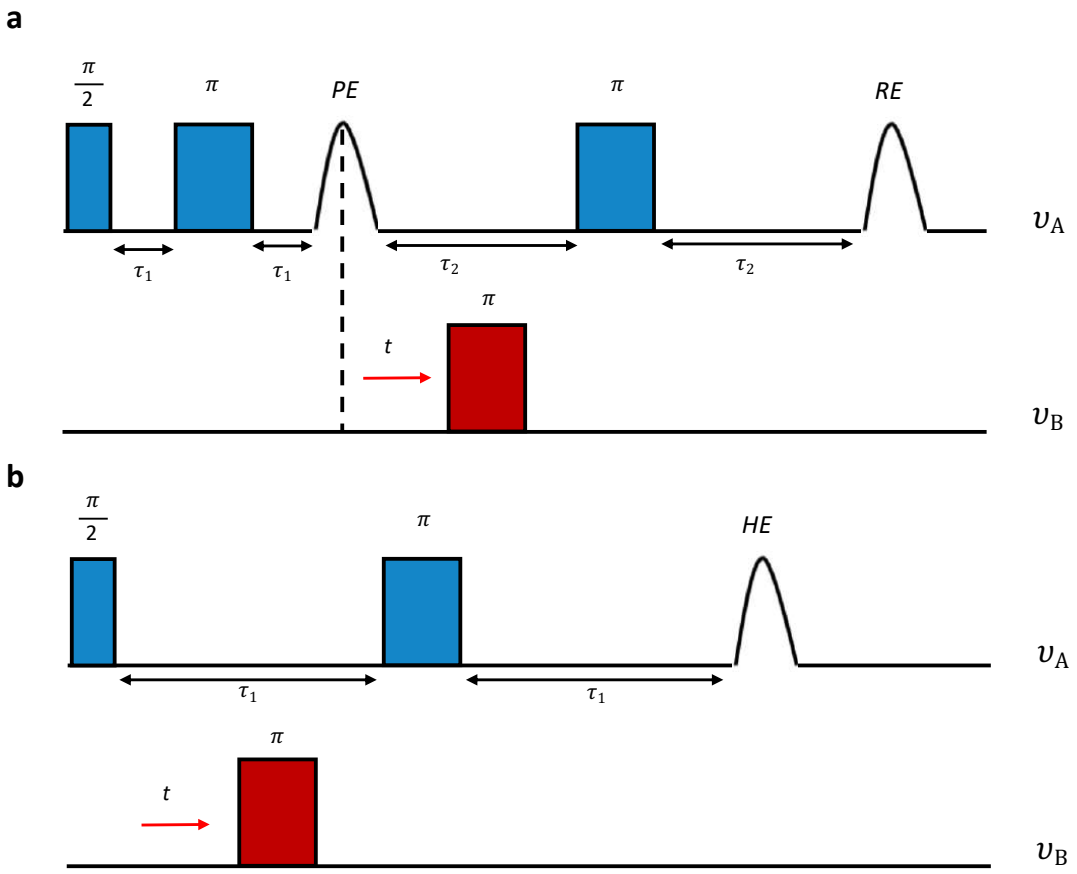


Figure 12. a) *4-pulse dead-time free PELDOR pulse sequence. Observer pulses at the first frequency are shown in blue, and the inversion pulse at the second frequency is shown in red. PE and RE correspond to primary echo and refocused echo, respectively.* ***b)*** *3-pulse PELDOR pulse sequence. Observer pulses at the first frequency are shown in blue, and the inversion pulse at the second frequency is shown in red. HE corresponds to a Hahn echo.*

Therefore, moving the inversion pulse in the time t, the echo intensity V(t) will oscillate according to the dipolar frequency ω_{dd}. The echo intensity can be expressed as:

$$V(t) = V_0 \cos(\omega_{dd} t) \qquad (8)$$

where V_0 is the echo intensity at t=0.

The five-pulse RIDME experiment[101] (**Figure 13 a**) is a dead-time free version of the three-pulse RIDME experiment[102], wherein akin to the four-pulse PELDOR experiment, an initial π/2 pulse is replaced by the sub-sequence (π/2-τ_1-π-τ_1) again forming a Hahn echo at time 2τ_1. The inversion π-pulse used in PELDOR is replaced by a 'mixing block' subsequence (π/2-T_{mix}-π/2), which causes longitudinal relaxation driven spontaneous spin flips of non-resonant 'pump-spins' so that $\Delta m_s \neq 0$. During the mixing block the A-spin magnetization is parallel to B_0, and the second π/2-pulse partially restores the A-

spin magnetization to the transverse plane. Analogously to PELDOR, a coherence transfer to resonant (detected) spins, induces a further contribution to their precession frequency by $\pm\omega_{dd}\Delta m_s$. The time of the mixing block is incremented over the series of performed experiments, still maintaining the dipolar evolution constant, before a final π-pulse at time $(t_2 - t)$ generates a refocused echo modulated by $\cos(\omega_{dd}t)$. This experimental set-up can be also renamed as "constant time" (ct) RIDME, since the total period of dipolar evolution (t_{max}) is identical (*i.e.*, constant) for all time-points through-out the measurement. This can limit the achievable sensitivity in dependence of the t_{max}, and simultaneously mandates selection of t_{max} a priori, optimized for a particular inter-spin distance. Moreover, an alternative RIDME pulse sequence which has a dynamic t_{max} and recently has been introduced, and has been called "variable time" (vt) RIDME[103] (**Figure 13 b**). In this sequence the t_{max} is incremented at each time-point throughout the measurement.

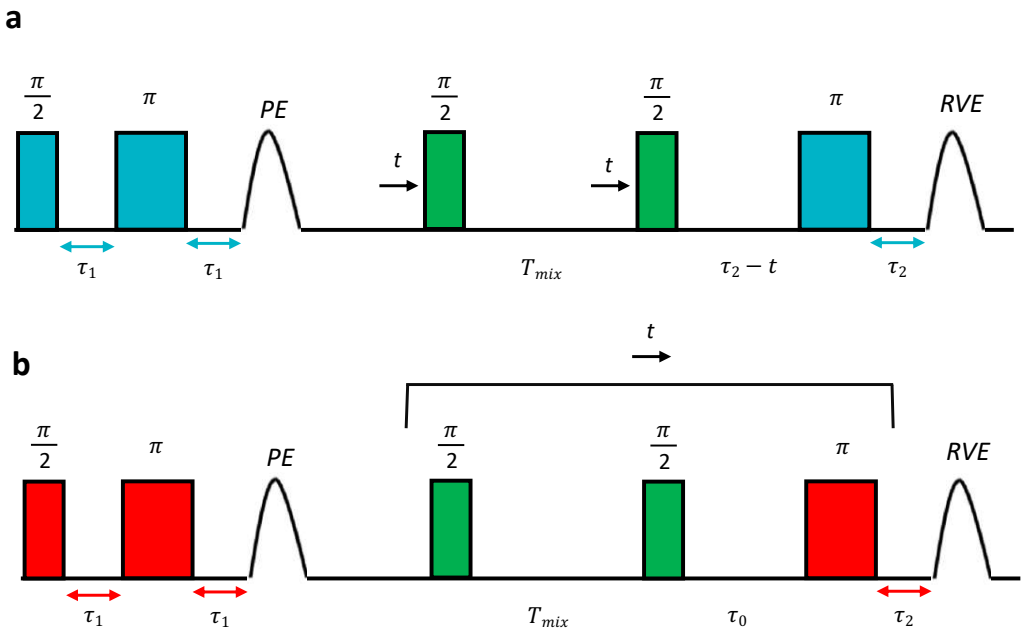


Figure 13. a) *5-pulse constant-time RIDME pulse sequence, with detection pulses indicated in blue, and mixing block interval (T_{mix}) pulses indicated in green. The position of the mixing block is incremented in time by t, indicated by the black arrows. The interval between detection pulses in the Hahn echo subsequence (τ_1) and the total dipolar evolution time (τ_2) are indicated by the double-headed blue arrows. The position of the final p-pulse is constant, separated by τ_2-t from the mixing block.* ***b)*** *5-pulse variable-time RIDME pulse sequence, with detection pulses indicated in red, and mixing block interval pulses in green. The intervals τ_1 and τ_2 are indicated by the double-headed red arrows. Here, the position of the mixing block and the final p-pulse are incremented together in time by t (i.e., the position of the final p-pulse is dynamic, separated by an arbitrarily short time interval (τ_0) from the mixing block.*

In this thesis both ct and vt RIDME were employed to systematically investigate two GB1 constructs with copper(II)-nitroxide orthogonal spin labelling (using four different nitroxide labels), alternating both labels between the α-helix and the β-sheet and their performances were compared.

RIDME is the ideal method for probing distance distributions between these distinct labels. This choice is driven by the inherent broadness of the copper(II) spectrum as a paramagnetic metal center, in contrast to the narrower spectrum of a nitroxide radical. Moreover, the two spectra are completely separated and do not overlap at Q-band EPR frequencies. Consequently, employing double-resonance techniques (such as PELODR), which need selective excitation of both spin centers, requires an extensive bandwidth, which is not readily available in most EPR resonators. Moreover, single frequency techniques that aim to excite both spins simultaneously (such as Double Quantum Coherence (DQC)) are even more challenging, given that they would need to cover both frequencies simultaneously. However, RIDME presents a solution by statistically inverting the fast-relaxing spin through longitudinal relaxation, thereby introducing the dipolar interaction. This approach means that the microwave pulses only need to excite a sufficient fraction of the nitroxide radical to detect the echo, a condition readily met in most experimental setups. The inter-spin distance can then be deduced by analyzing the modulation with the dipolar frequency. An additional advantage of RIDME is that it circumvents orientation-dependent excitation of the broad copper(II) spectrum, simplifying the analysis and interpretation process as all orientations contribute.

4.2.3.3. RIDME measurements experimental set up

All RIDME experiments, for both the constant time (ct) and the variable time (vt), were carried out at 30 K. For every sample, the pulses were applied on the maximum of the nitroxide echo detected field sweep. ctRIDME measurements were performed using the pulse sequence $\pi/2–\tau_1–\pi–(\tau_1+t)–\pi/2–T_{mix}–\pi/2–(\tau_2-t)–\pi–\tau_2–$echo with detection pulse lengths $\pi/2$ and π, respectively of 8 and 16 ns. Each trace was acquired using an SRT of 10 ms, a τ_1 of 400 ns, and 2 shots-per-loop and 32-steps phase cycling.

vtRIDME measurements were performed using the pulse sequence $\pi/2-\tau_1-\pi-(\tau_1+t)-\pi/2-T_{mix}-\pi/2-\tau_0-\pi-(\tau_2+t)-$echo with detection pulse lengths $\pi/2$ and π, respectively of 8 and 16 ns. Each trace was acquired using an SRT of 10 ms, a τ1 of 400 ns, and 2 shots-per-loop and 32-steps phase cycling.

Measurements were recorded with a short (reference) and a long mixing time of 5 and 200 μs, respectively, to allow deconvolution (dividing the constant and variable time RIDME traces with the longer mixing times by the corresponding reference traces) of the traces.

4.2.3.4. Background correction

The detected signal of both the PELDOR and RIDME experiments (and, in general, all pulse dipolar EPR methods) has contributions that derive from the intramolecular couplings of electron spin pairs, V_{intra}, and an intermolecular component from couplings of electron spins on discrete macromolecules, V_{inter} as expressed below:

$$V(t) = V_{intra}(t)\, V_{inter}(t) \tag{9}$$

Since the dipole-dipole coupling frequency information is encoded only by V_{intra} the intermolecular component needs to be extracted from the overall experimental data, so to obtain the intramolecular component *i.e.* that of interest. In the PELDOR experiment, this intermolecular signal component is approximated as a stretched exponential background function. On the other hand, the intermolecular decay in the detected RIDME signal is dominated by electron-electron and electron-nuclear spectral diffusion. These spectral diffusion processes manifest largely during the mixing block interval since it is often on the μs timescale, and result in RIDME background decay often being considerably steeper than PELDOR traces measured with the same dipolar evolution period. Since a full derivation of the RIDME background function is beyond the scope of this thesis, it is sufficient to describe it as a stretched exponential with dimensions between three and six.

4.2.3.5. Data Analysis of dipolar traces

The DeerAnalysis[73] MATLAB program is among the most used softwares for extracting distance distributions from dipolar traces. This software allows for a different selection of background choices, from homogeneous background correction (which is equivalent to an exponential background signal), to different degrees of polynomial background correction. At the same time, DeerAnalysis allows to retrieve distance distributions P(r) using a numerical algorithm called Tikhonov Regularization. Converting the dipolar evolution into a distribution of distances P(r) is an ill-posed problem, since small changes in input data, for instance due to small errors in the separation of the desired trace and the background, can lead to large changes in the output distance distribution. Considering a known distance distribution P(r) the corresponding trace can be computed by multiplication with a kernel function K(t,r):

$$\mathrm{V}_{intra,sim}(t) = \mathrm{K}(t,\, r)\ \mathrm{P}(r) \tag{10}$$

This ill-posed problem can be overcome using the Tikhonov regularization approach, which is able to fit the best distance distribution from the experimental data by minimizing the differences to achieve the optimal compromise between the artefact suppression and the resolution in the distance distribution.

$$G_\alpha = \left\| V_{intra,\,simulated}(t) - V_{intra}(t) \right\|_\alpha^2 + \alpha \cdot \boldsymbol{\eta}(\boldsymbol{\alpha}) \tag{11}$$

$$\eta(\alpha) = \left\| \frac{d^2}{dr^2} P(r) \right\|_\alpha^2 \tag{12}$$

While first term represents the deviation between the experimental data $V_{intra}(t)$ and the theoretical simulated function $V_{intra,simulated}(t)$, called $\rho(\alpha)$, the second term stands for the product between the regularization parameter, α, and the fit smoothness $\eta(\alpha)$, which is defined as the norm of the second derivative of the distance distribution. In this thesis DeerAnalysis was widely employed for extracting the distance distributions from several RIDME dipolar traces.

4.2.4. Site Direct Spin Labelling in EPR

Among the vast number of existing biomolecular systems, only a small percentage can be directly subjected to continuous wave and pulsed EPR since only a few of them contain intrinsic paramagnetic moieties, such as metalloenzymes or iron-sulfur (Fe-S) proteins. Therefore, a significant challenge arises when studying biomolecules that lack an inherent paramagnetic center, such as DNA, RNA, or certain proteins. The following section aims to explain how successfully introduce a paramagnetic spin probe through Site Direct Spin Labelling (SDSL).

4.2.4.1. The SDSL reaction

As already introduced, SDSL is nowadays considered the cornerstone technique for selectively introducing small paramagnetic labels, mostly nitroxides, into specific and strategically placed sites within otherwise diamagnetic proteins. Briefly, this approach focuses on introducing a cysteine residue (if not already present) on the desired position, through site direct mutagenesis, while mutating all the other unwanted cysteines with different amino acids (*e.g.*: alanine, serine, histidine). The stable nitroxide spin label is then covalently grafted to the protein surface through the formation of a disulfide bond with the cysteine residue (**Figure 14a**).

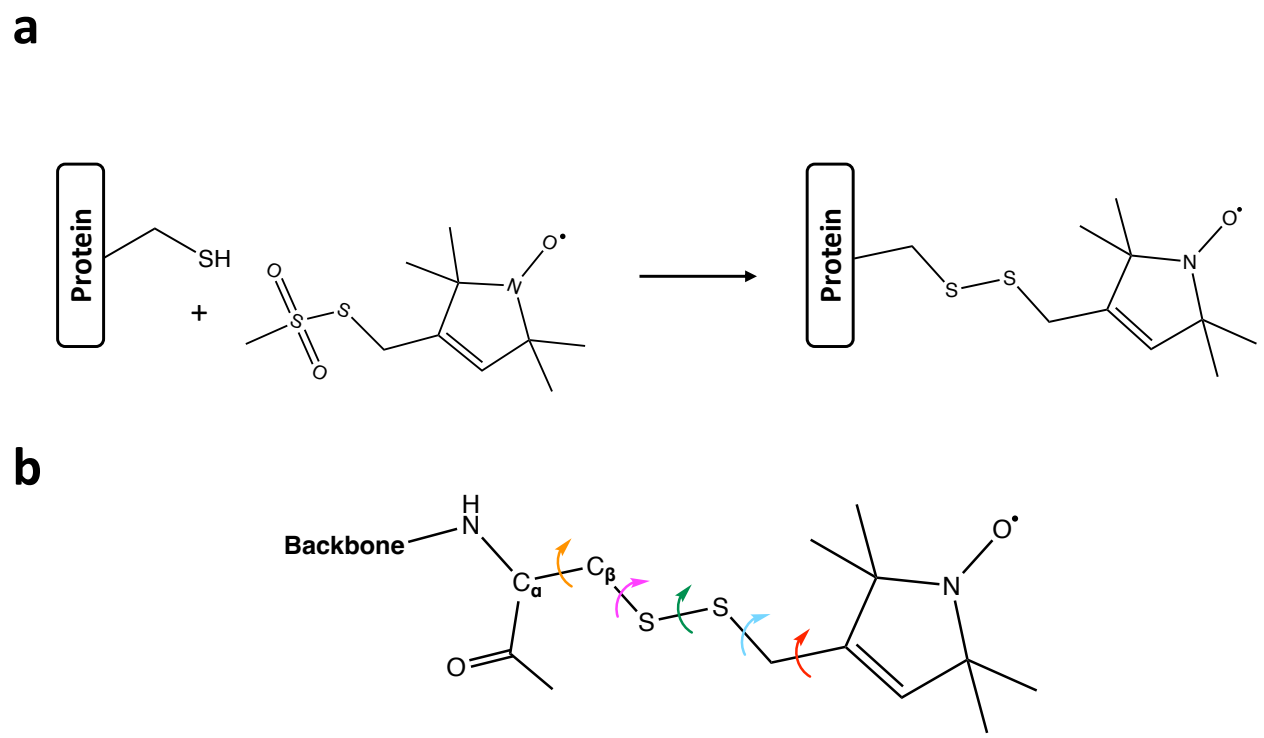


Figure 14. a) *SDSL reaction between a 1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTSL) nitroxide spin label and a cysteine residue on the protein backbone with consequent formation of the disulfide bond.* ***b)*** *Detailed MTSL structure highlighting the five torsion angles (red, light blue, green, magenta, and orange).*

Once the label is successfully attached to the site of choice, insights about its dynamics can be directly retrieved from the cw-EPR spectrum. In general, the spin label choice is crucial when performing EPR experiments, since distance measurements can be heavily affected by the intrinsic features of the selected probe, such as the rigidity, the linker length and the sensitivity. The next section is meant to introduce the most exploited spin label classes in the context of EPR.

4.2.4.2. Spin Label Classes

The most important property of an ideal spin label is undoubtedly its intrinsic chemical stability. The label should be as small as possible, to perturb the protein structure as little as possible. It should react in mild conditions and contain a reactive group that allows incorporation into the protein with high selectivity towards a particular amino acids type. All spin labels can be divided in three main groups based on the spectroscopic properties of the paramagnetic moieties. The first group is represented by the metal/bound complexes with metal ions ranging from lanthanides ions, such as gadolinium (Gd(III))[104]

to transition metal ions such as copper (Cu(II))[23,105] and manganese (Mn(II))[106,107]. The second class is related to triarylmethyl (TAM)[108] radicals, (*e.g.*: SLIM radical[109]), that are often preferred to the nitroxide radicals for in-cell applications due to their higher stability in the intracellular reductive environment. The last class and, as of date, the most exploited one is the nitroxide radicals.

4.2.4.2.1. Nitroxide Spin Labels

Based on the numbers of carbon atoms present in the ring containing the nitroxide moiety we can distinguish two different subgroups of spin labels: the five member rings, such as MTSL, and the six member rings, like the 2,2,6,6-tetramethyl-N-oxyl-4-amino-4-carboxylic acid (TOAC)[110]. The four tested labels were: the already introduced MTSL (or MTSSL)[62,111], 3-Maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (MPSL or malemide-Proxyl)[112-114], 3-(2-Iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (IPSL)[115-117] and bis-(2,2,5,5-Tetramethyl-3-imidazoline-1-oxyl-4-yl)disulfide (IDSL)[118-121] (Errore. L'origine riferimento non è stata trovata.).

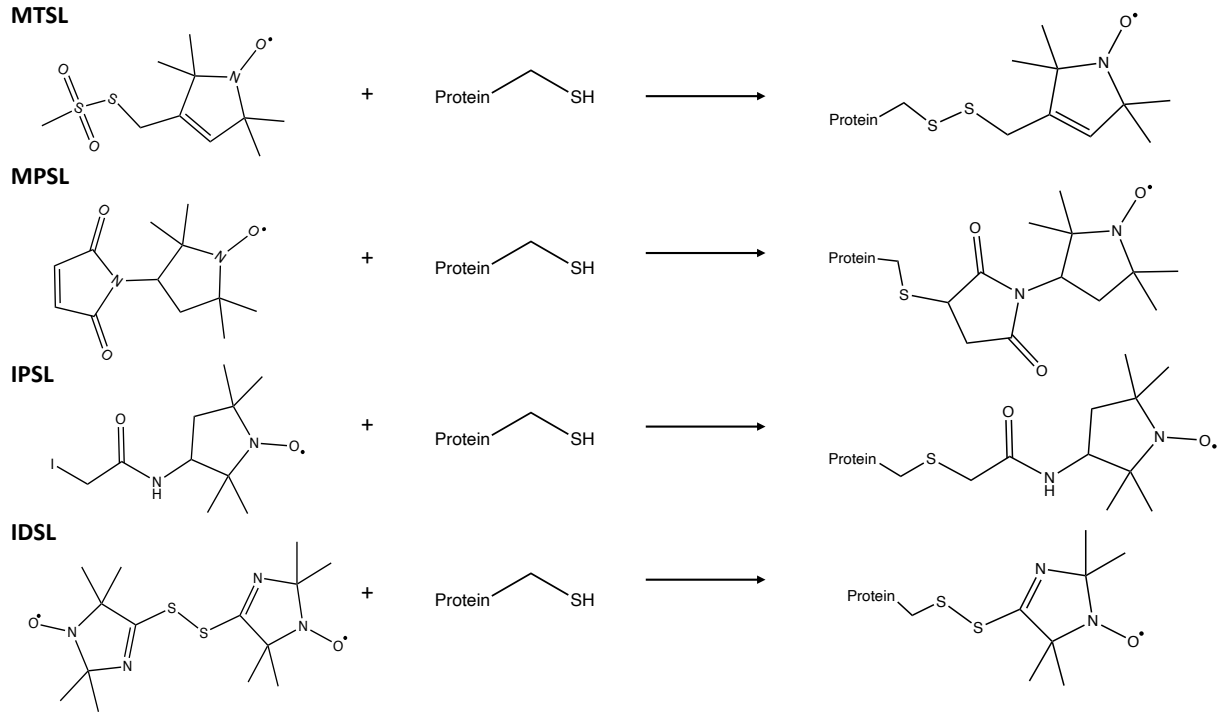


Figure 15. *Chemical structure of the four different nitroxides investigated. Top left, 1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTSL); Bottom left. 3-Maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (MPSL); Top right, 3-(2-Iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (IPSL); Bottom right, bis-(2,2,5,5-Tetramethyl-3-imidazoline-1-oxyl-4-yl)disulfide (IDSL). The reaction of each label with the cysteine residue is reports as well.*

As already described, the MTSL is the "workhorse" among all nitroxide spin labels, however the disulfide bond between the cysteine and the labels itself can be easily cleaved in reductive conditions, such as intracellular environment. Meanwhile, the presence of the malemide moiety (*e.g.*: MPSL) creates a carbon-sulfur bond between the label and the cysteine that is less likely to be reduced, making it a better candidate for in-cell studies (further details are discussed in Section 4.3.4). The iodoacetamido (IPSL) is another good option for potential in-cell studies since it reacts with the thiol group forming a novel C-S bond. However, its reactivity towards the -SH group seems reduced respect to MTSL and MPSL, and during the labelling reaction a great excess of nitroxide reagent is needed to obtain high reaction yield[122]. Lastly, the IDSL is a biradical spin label, with the shortest linker between the cysteine residue and the unpaired electron. It is not easily solubilized in aqueous buffer and often sonication cycles must be carried out to achieve full dissolution. It reacts by substituting one sulfide of its disulfide bond with the one of the cysteine thiol groups. In contrast to substitutions with good leaving groups (MTSL and IPSL) or addition reactions (MPSL), this disulfide exchange has an equilibrium constant closer to unity, thus incomplete labelling and free dimeric label are very likely[122].

4.2.4.2.2. Cu(II) Spin Labels and the Double Histidine Motif

Metal-ion based labels are a valuable alternative to nitroxide radicals. In particular, Cu^{2+} ions bound directly to native binding sites in the protein have proven to be a significant choice for structural determination of proteins[123-125]. However, when proteins lack intrinsic Cu^{2+}-binding sites, alternative spin labelling techniques have been developed to introduce Cu^{2+} site-specifically into a protein using chelating tags. Initially, these tags, such as the 1-(2-(pyridin-2-yldisulfanyl)ethyl)-1,4,7,10-tetraazacyclododecane (TETAC), were employed for cysteine conjugation and they showed high affinity for chelating Cu^{2+} ions. However, their high flexibility provides distributions with a width in the range of the one of the MTSL tag[126]. An alternative labelling technique for binding Cu^{2+} involves the use of two strategically placed histidine residues that create the double histidine (dHis) binding motif[23] (already introduced in section 2.2) (**Figure 16**). This motif involves placing histidine residues at positions *i, i + 4* if positioned on a α-helix and *i, i + 2* for a β-sheet. Despite the need to perform a double mutation for each dHis-Cu^{2+} label, this motif

produces distance distributions almost five times narrower than the one obtained from MTSL. This narrowness derives from the higher intrinsic rigidity of the dHis-Cu^{2+} system. However, at the beginning, a big limitation on the use of these kind of systems derived from the poor selectivity of the Cu^{2+} towards the dHis motif, which may coordinate nonspecifically to other residues on the protein. To address this issue, Cu^{2+} ions were chelated by ligands such as iminodiacetic acid (IDA)[23,127] or nitrilotriacetic acid (NTA)[128], before the actual reaction with the dHis motif (**Figure 16**).

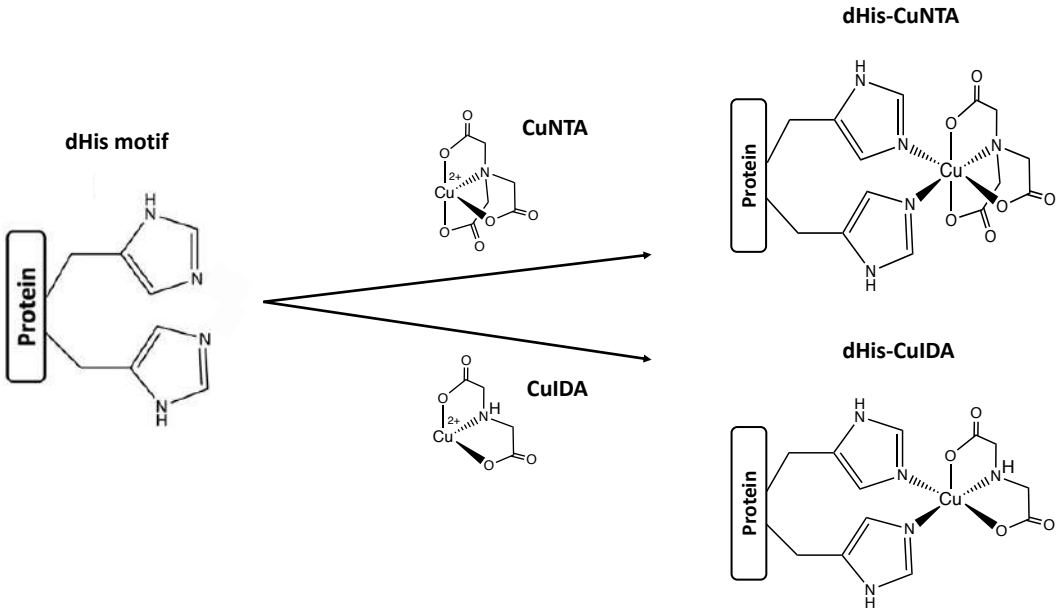


Figure 16. *on the left, schematic representation of the double histidine motif (dHis) and the chemical structure of the copper nitrilotriacetic acid (CuNTA) and copper iminodiacetic acid (CuIDA). On the right, the structure of the double histidine motif after binding the two* Cu^{2+} *chelator agents, resulting respectively in the dHis-CuNTA and the dHis-CuIDA.*

Alternative residues to cysteine, as selected sites for paramagnetic labelling, are often required for spin labelling. They can be tyrosines[58], but above all histidine residues. Using CuNTA or CuIDA increases the binding specificity of the complex towards the dHis, while unspecific binding to the other protein sites is reduced. As discussed, the double histidine motif, being highly rigid, provides narrower distance distribution, in respect to the nitroxide-nitroxide case. This feature is particularly appealing because spin labels side chains with a considerable degree of flexibility do not provide clear insights on whether the width of the distance distribution depends on the intrinsic motions of the spin labels or on the actual protein backbone fluctuations[129].

4.2.5. *In silico* approaches

The possibility to simulate distance distributions between paramagnetic moieties can be fundamental for gathering valuable insights into the investigated system before relying on experimental approaches. By modeling the spin label conformers through *in silico* approaches it is possible to identify suitable labelling sites by performing a systematic scanning of the protein residues. The two main *in silico* modelling tools used in this project were MtsslWizard[130] and Multiscale Modeling of Macromolecules (MMM)[131]: both provide rotamers assemblies of spin labels that are then used to derive distance distribution, but are based on different theoretical backgrounds. The following sections give a brief explanation on their functioning.

4.2.5.1. Multiscale Modeling of Macromolecules

MMM is an open-source toolbox implemented in MATLAB, with a graphical user interface, which allows to simulate spin label rotamers conformation and the dipolar trace and distance distributions. It works by exploiting a library of rotamers based on the conformational space of the spin label. The set of sterically unrestricted rotamers of the spin label is precomputed in advance, with an estimate of their relative free energy ($f_{o,i}$). Afterwards, all rotamers *i* are sequentially simulated and attached *in silico*. Their interaction energy (Δu_i) with the macromolecule is calculated, considering only repulsion and van der Waals interactions via a Lennard-Jones potential. Lastly, populations (p_i) for each rotamer are then calculated using a Boltzmann distribution with $f_{o,i} - \Delta u_i$ as an estimate for their free energies. The electron spin spatial distribution is represented by a cloud of spin density centers from all rotamers, considering the populations (p_i)[131]. Beyond the possibility to simulate several nitroxide radicals, recently, a new development allows to model Cu(II) chelator agents as spin labels for the dHis motif[132]. In this thesis this new feature was exploited to model distance distribution between the nitroxide and CuNTA spin labels. Modelling was performed both under ambient (298 K) and cryogenic (175 K) temperatures (**Figure 17**).

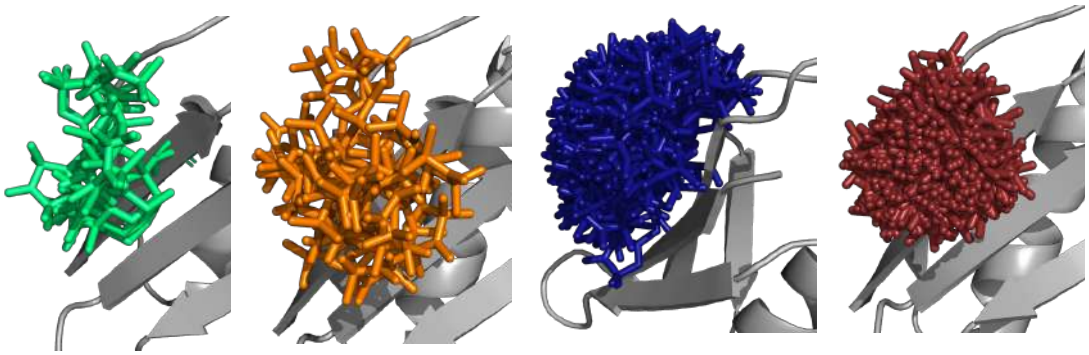


Figure 17. *Representation of all MTSL modelled rotamers for the two in silico labelling approaches, MMM with cryogenic (green rotamers) and ambient (orange rotamers) conditions, and MtsslWizard with Tight (red rotamers) and Loose (blue rotamers) settings.*

4.2.5.2. MtsslWizard

Introduced in 2012 as a Pymol plugin-in[130], MtsslWizard follows the "tether-in-a-cone" approach toward the location of the unpaired electron on the spin label. This approach assumes that the label must be situated somewhere within the conical area surrounding the attachment site and predicts where on the cone the label will most likely be situated, screening only for rotamers that do not clash with the protein structure. The software allows to select the number of hypothetical clashes between the protein and the modelled rotamers. Even in this case, recent development allowed us to simulate both rotamers conformers and distance distributions between nitroxides and CuNTA spin labels[133]. In this thesis we worked with two different settings. The first one named as "Tight", with vdW-restraint 0 clashes and 3.4 Å cutoff (where a clash is defined as a violation of a "vdW cutoff" between label and protein), while the second one, "Loose" with vdW-restraint 15 clashes and 2.0 Å cutoff (**Figure 17**). The Loose setting allows closer contact between the label and the protein, while Tight settings reduces the proximity between the label and the protein, imposing a more conservative distance threshold.

4.2.6. Structure Predictor Tools

In late 2021, DeepMind made a major breakthrough in the decades-long quest to find a way to accurately predict protein structures without the need for experimental techniques[134], such as X-ray crystallography, NMR, and cryo-EM. Indeed, these experimental methodologies can be very time-consuming and not always applicable to every system under investigation. Instead, they successfully managed to create a

revolutionary approach, AlphaFold2, which computationally predicts reliable protein structures. This neural-network-based model was trained on more than 170000 proteins structures taken directly from the worldwide Protein Data Bank (wwPDB), becoming the first of its kind to predict with great accuracy three-dimensional multimeric protein structures. This revolutionary approach greatly broadened the applicability of computational methodologies in predicting protein structures in a reliable way. Onwards, the introduction of AlphaFold2 has inspired the development of other structural predictors tools such as OmegaFold[135], ESMFold[136] and RoseTTaFold[137]. In the sections below AlphaFold2, OmegaFold and ESMFold will be discussed more in details since all three methods were employed for the prediction of GB1 structures exploited in the paper reported at the end of the chapter.

4.2.6.1. AlphaFold2

AlphaFold2 is an attention-based neural network used in protein structure prediction. It is based on using large datasets of protein sequences with known structures to construct multiple sequence alignments (MSAs) within the network. This allows the model to identify similarities between the protein sequence and those in the known database. Afterwards, the system forms an initial representation of the structure (the 'pair representation'), giving an early estimate of which amino acids are likely to be in contact with one another. Information is then exchanged between the MSAs and the pair representations, following 48 repeated iterations before finally both the MSAs and the pair representations converge into a stable 3D structure prediction of the protein[134] (**Figure 18**).

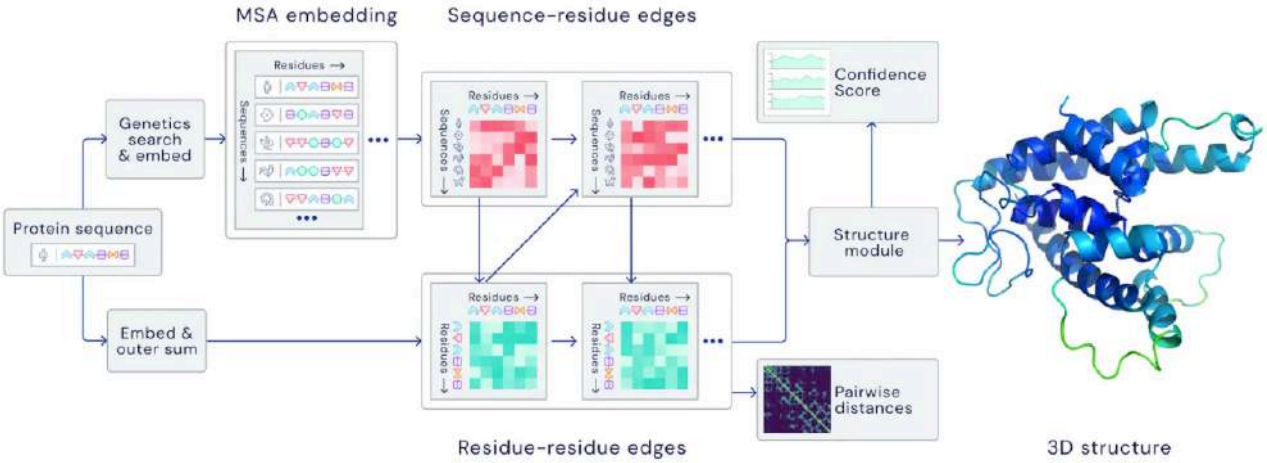


Figure 18. *Schematic workflow of neural-network system AlphaFOld2 taken from DeepMind blogpost in November 2020.*

4.2.6.2. OmegaFold

OmegaFold uses a deep-learning model to predict high-resolution protein structures from a single primary sequence without relying on MSAs, unlike AlphaFold2. The process begins with a large pretrained language model for sequence modeling, called OmegaPLM. This model is meticulously trained on an extensive dataset of unaligned and unlabeled protein sequences, enabling it to learn single and pairwise residue embeddings, capturing crucial structural and functional information embedded within amino acid sequences. These embeddings are then passed to Geoformer, a geometry-based transformer neural network, whose primary role is to ensure the geometrical consistency of the embeddings by ensuring that amino acid node and pairwise embeddings generate coherent coordinates and distance predictions when projected into 3D space. Lastly, the information from OmegaPLM and Geoformer is harnessed by a structural module to predict the three-dimensional coordinates of all heavy atoms into the protein[135] (**Figure 19**).

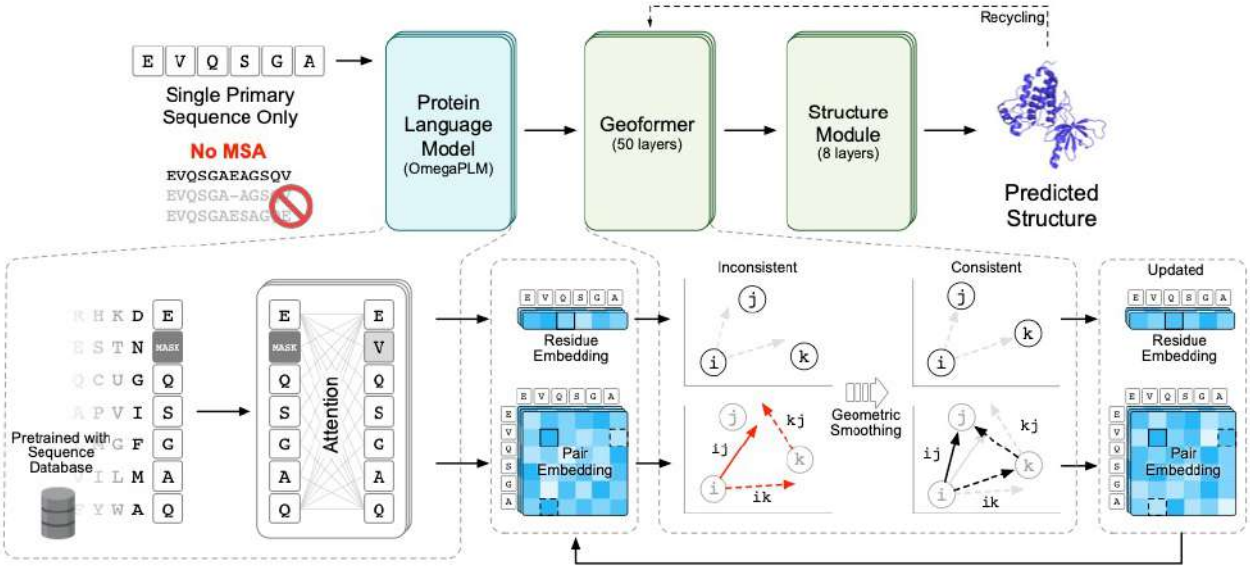


Figure 19. *OmegaFold schematic workflow, image taken form the 2022 introductory paper on Omegafold*[135].

4.2.6.3. Evolutionary Scale Modeling Fold

Like Omegafold, ESMFold works by using a language model to generate a three-dimensional structure of a protein directly from its primary amino acid sequence, without relying on MSA. ESMFold works by passing the protein sequence through a pretrained language model (ESM-2), which generates representations of the sequence that capture the evolutionary patterns linked to structure. These representations are then passed to a folding head, which predicts the 3D structure of the protein. The folding head is a neural network that consists in a series of folding blocks, and each folding block alternates between updating a sequence representation and a pairwise representation. The output of the folding blocks is passed to an equivariant transformer structure module, which predicts the 3D structure of the protein[136] (**Figure 20**).

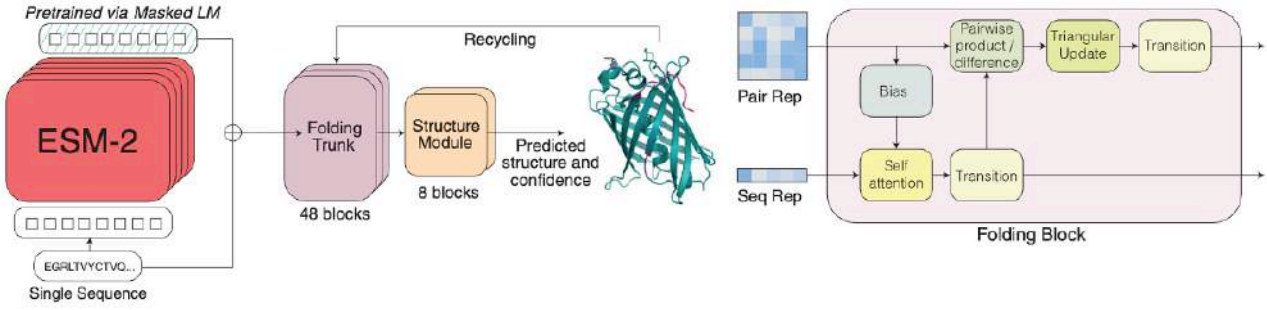


Figure 20. *ESMFold schematic workflow, image taken form the 2023 introductory paper on ESMFold*[136].

4.3. In-cell applications

This section briefly introduces the in-cell EPR spectroscopy approach explaining its advantages over the *in vitro* approach in investigating proteins directly inside their intracellular environment. The basic idea behind employing in-cell based experiment is to investigate proteins affected by their physiological environment, which could influence their structure and dynamics resulting different with what they have *in vitro*. In this thesis the heat shock delivery method was exploited to introduce protein both in prokaryotic (*Escherichia Coli)* and eukaryotic (*Pichia Pastoris)* cells. The most popular techniques for obtaining in-cell samples are briefly described as well.

4.3.1. The in-cell approach

In-cell structural biology aims to uncover the structural, functional, and dynamic information of biomacromolecules, such as proteins and nucleic acids, within the context of living systems. Key players in this pursuit are Electron Paramagnetic Spectroscopy (EPR) and Nuclear Magnetic Resonance (NMR), which provide invaluable insights into molecular structures and behaviours. Traditionally, structural biology has always related on the robust *in vitro* approach, which aims to investigate the systems in an environment as close as possible to the biological one, carrying out all experiments in stable water-based buffer solution. However, this method does not fully capture all the features of biological systems within their natural intracellular milieu, which encompasses the cytoplasm and various intracellular organelles. As a result, there is a growing desire to obtain a more precise depiction of proteins within their native environment, which is now becoming achievable thanks to recent advancements[138,139]. In the early 2000s, for the first time, in-cell magnetic resonance experiments were performed allowing to characterize how intracellular crowding, local pH changes, non-specific and specific interactions occurring inside the cellular compartments, affect the investigated biomacromolecules[140-143].

4.3.2. Spectroscopical methods for in-cell

Both in-cell NMR and EPR have been employed for a wide range of applications on insect[144], bacterial[145-147], yeast[148] and human cells[32,149-151]. In-cell EPR/pulsed EPR spectra

have minimal interference from background signals and therefore can be essentially considered as background-free. CW-EPR is extremely useful in reporting the information about the intracellular local dynamics of the paramagnetic tags, mostly nitroxide radicals, attached to the biomolecules systems. In fact, as already discussed at the beginning of the thesis, the spectral line shapes are highly affected by intracellular environment which significantly affects the overall correlation time of the paramagnetic spin probe[4,152]. On the other hand, pulse dipolar EPR techniques are able to detect the changes on the distance distributions of several proteins between the *in vitro* and the in-cell samples[85,89,90]. Therefore, EPR is a powerful and attractive tool for in-cell investigations, although careful considerations, especially regarding the selection of the appropriate label for in-cell experiments (see Section 4.3.4), are necessary to optimize its applications.

4.3.3. Methods for sample preparations

Few steps need to be carried out to correctly prepare a sample for in-cell EPR or NMR investigation. Regarding EPR, the investigated biomacromolecules should contain a suitable paramagnetic label into the appropriate position exploiting the SDSL approach. The second step concerns the actual introduction of the investigated protein inside the intracellular environment. Similarly, to what already discussed for the ^{19}F labelling, two main categories of approaches - overexpression and delivery-based methods - are exploited for in-cell sample preparation. In the next section the delivery-based approaches will be briefly introduced since one of this methodology has been exploited in the context of this thesis.

4.3.3.1. Delivery-based methods

This category of techniques is based on the delivery of the already expressed, purified, and modified proteins into the intracellular environment of different living systems. A schematic representation of the methods that will be discussed in this chapter are reported in

Figure 21.

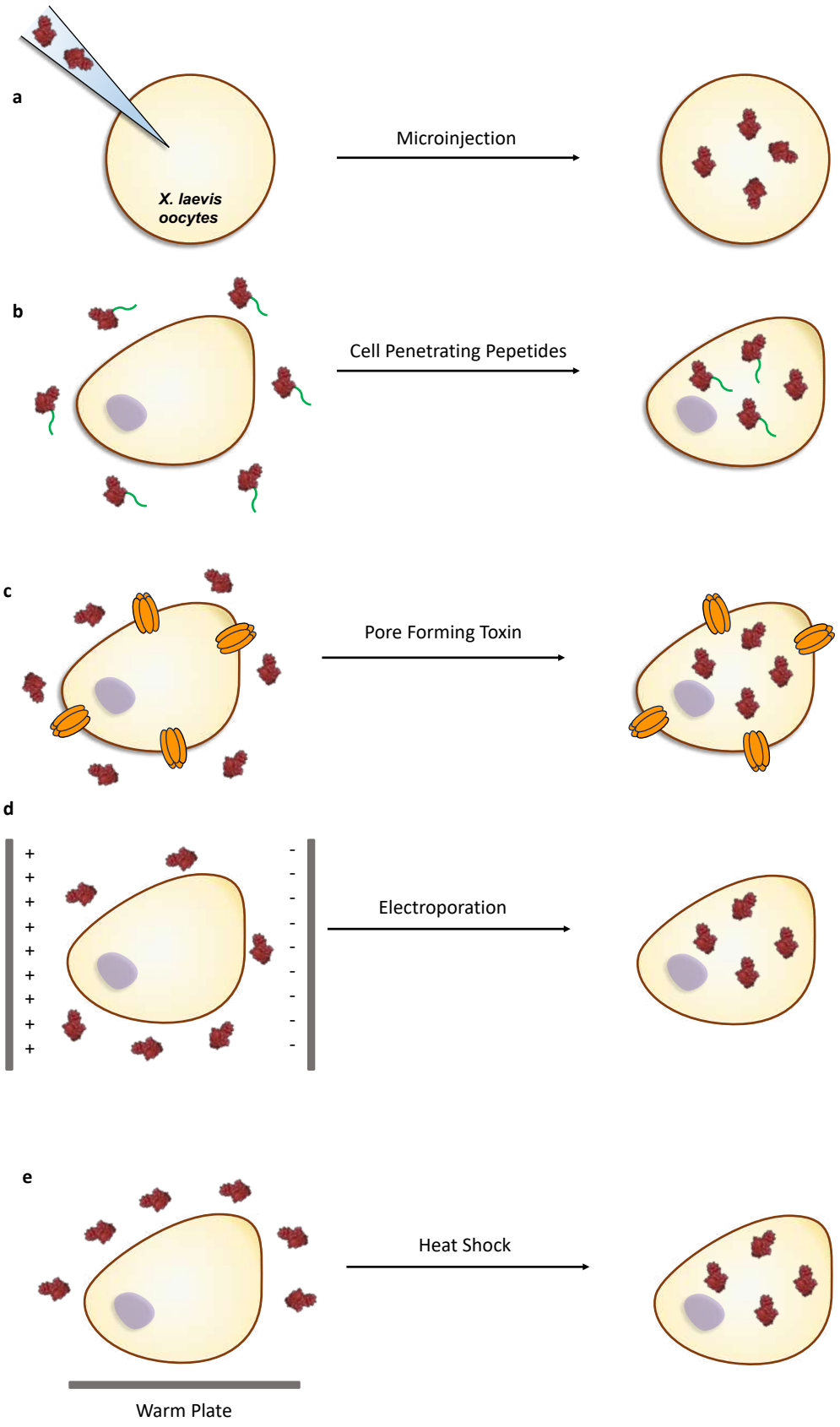

a
X. laevis
oocytes
Microinjection
b
Cell Penetrating Pepetides
c
Pore Forming Toxin
d
+
+
+
+
+
+
+
+
+
-
-
-
-
-
-
-
-
-
Electroporation
e
Heat Shock
Warm Plate

Figure 21. *Schematic representation of protein (in red) delivery-based methods for the preparation of in-cell samples.* ***a)*** *Mechanical delivery based on microinjection of the protein frog/fish oocytes.* ***b)*** *Delivery exploiting Cell Penetrating Peptides (CPP) covalently linked to the proteins (in green) to help the proteins through the cell membrane.* ***c)*** *Pore Forming Toxin delivery approach which through oligomerization of PFTs creates pores (in orange) that allow the delivery of protein.* ***d)*** *Electroporation delivery where a voltage is applied to a suspension of cells and protein. The electric pulses stimulate the formation of the pores (in orange), which are used from the protein to enter the intracellular environment.* ***e)*** *Thermal stimulation delivery method, where a solution of cells and proteins are heated and then incubated in ice. The methods b, c, d and e have been successfully employed both with bacterial and yeast cells, but also with human cells.*

The mechanical microinjection approach was one of the first delivery approaches to be exploited and it is still greatly used nowadays[153-155]. It requires the use of big enough cells, such as *Xenopus Laevis* oocytes, to manually introduce, through a microinjector, the desired biomolecules. Another delivery-based approach is the Cell Penetrating Peptides (CPP) method. It was first introduced in the context of in-cell NMR where short peptides, around 10-30 amino acids, with the ability to cross the cell membrane, are covalently linked to the protein[141]. Once inside the intracellular environment the CPP flags are released from the proteins by endogenous enzymatic activity or by autonomous reductive cleavage, allowing the protein to freely interact with endogenous biomolecules. So far, this method has yet to be employed in combination with in-cell pulsed EPR; however, a recent work successfully tracked the delivery of a nitroxide radical directly inside human cell with cw-EPR by exploiting the recombinant analogue of human k-casein fragment to the nitroxide[156]. The Pore Forming Toxins (PFTs) is another delivery approach mostly applied for in-cell NMR studies[157], but with some applications even in the EPR field[158]. Pore Forming Toxins are a class of proteins located on membranes that can form, through an oligomerization process, 20-50 nm wide pores, which allows the delivery of biomolecules form the external buffer to the intracellular environment[159]. After the pores are assembled and the protein is delivered, their resealing is achieved upon the supplementation of Calcium ions (Ca^{2+}) in the cytosol[160], limiting this approach only to the cells that are able to properly reseal. Among all the delivery methods, electroporation[161] was initially developed for delivering heterologous DNA into different living systems and only subsequently expanded to protein delivery[162]. This method involves the application of electric pulses of 1kV/cm in hundreds of microseconds with the aim of achieving permeabilization of the cell membrane by forming pores up to 50 nm. This approach has been successfully applied for delivering different kind of proteins,

from globular to intrinsically disordered (IDPs) ones, with applications for both in-cell NMR[163,164] and EPR[85,162,165,166].

The delivery method that I have employed in my work is based on the heat shock approach, which was already shown to be successful for both bacterial and yeast cell[147], and has been implemented for human cells usage[151]. As electroporation, this simple, fast and efficient method has been originally developed for delivering nucleic acids into bacterial cells[167]. In this approach vesiculation is induced by a heat-shock cycle at (42 °C) for a defined period of time (usually 45 to 60 s). The temperature is close to physiological conditions to avoid underside folding issues.

4.3.4. Spin labels for in-cell

As already mentioned in Section 4.2.4.1.1, not every spin label is suitable for in-cell studies. In cellular context, the ideal spin-label should not only preserve the activity and folding of the biomolecule on which it is conjugated but should also be stable towards bond cleavage (between the label and the protein) or quenching (of the unpaired electron) by the cellular environment. Therefore, the natural reducing intracellular environment, crucial for the cell metabolism, does not allow the use spin labels that attach cysteine residues through the formation of sulfur-sulfur bonds (such as MTSL), because they are rapidly released as free labels. Hence, for in-cell investigations, spin labels forming a more stable carbon-sulfur (C-S) bond[109,154,168,169], like MPSL, are preferred. Moreover, nitroxide radicals can participate also to intracellular redox processed, leading to a conjugated diamagnetic EPR silent hydroxylamine analogue, with a consequent loss of EPR signal (**Figure 22**).

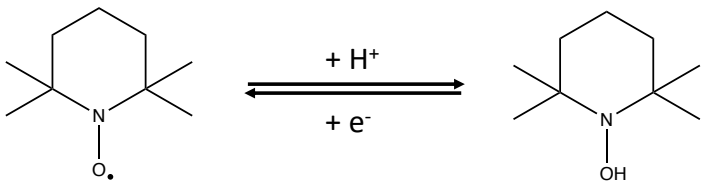


Figure 22. *Oxidation and reduction reaction of a nitroxide radical.*

While these labels have found minor applications in tracking cellular metabolism, their utility for structural studies within the cell at room temperature (cw-EPR) is impaired by the loss of the EPR signal over time due to reduction[170]. Consequently, reduction resistant spin labels, such as metal-based labels, or trityl radicals have been favored as the choice of spin labels for in-cell applications[85,109,168]. However, extensive efforts have been carried out to enhance the stability of nitroxide based spin labels in reducing environments. Their stability can be easily assessed *in vitro* by monitoring the disappearance of the EPR signal in the presence of reducing agents such as ascorbic acid, DTT (dithiothreitol) or TCEP (tris(2-carboxyethyl)phosphinehydrochloride). In general, it results that five-membered ring nitroxides tend to be more stable towards reduction compared to six-membered ones, with bulky substitutions in the α-position further enhancing stability[169]. Notably, replacing the four methyl groups in gem-dimethyl nitroxide with ethyl groups has shown to be the most effective substitution. Moreover, recently, increased stability of nitroxide radicals in reductive environments has been also obtained by the addition of the maleimide reagent in-cell lysate, which, by consuming NADPH/NADH (essential cofactors for the enzymes involved in nitroxide reduction), can stabilize nitroxide spin labels[171].

Among the metal-based EPR spin labels, gadolinium (Gd^{3+}) probes offer complete stability against reduction events[90,166,172]. Nevertheless, their application is challenged by their size and, notably, the length of their linker, which can affect the extracted distance distribution. Ongoing research, particularly led by Goldfarb group, continues to refine Gd^{3+} based labels to mitigate these challenges[89,166,172]. However, in general, the stability of the nitroxide labels is closely linked to the overall time required for in-cell sample preparation. For instance, in the case of the MPSL radical exploited in this thesis, a fast protein delivery protocol allowed for efficient cw-EPR data acquisition.

4.3.5. Deuteration in EPR

As already discussed, distance measurements by pulse EPR techniques, are extremely useful in the range between 2 and 10 nm, and sometimes even beyond[99,173]. At the long distance extreme of the measurable distances, the frequency becomes so small that it is challenging to capture enough time data to measure unambiguously the dipolar frequency. To measure increasingly longer distances, the major limiting factor is the

relaxation time T_2 of the signal, also addressed as T_m (phase memory time), since distance measurements are carried in frozen solution, which describes the nonrecoverable dephasing of the electron spin echo (ESE) in the xy plane. It is widely accepted that interactions with the protons of the solvent, typically water, highly contribute to the dephasing seen in spin-labeled protein solutions. Due to the deuteron smaller magnetic moment, (dipolar coupling is proportional to the magnetic moment), replacing the water with deuterated water (D_2O) caused a relatively high reduction in the rate of electron xy dephasing. Moreover, exploiting a deuterated cryoprotectant, such as glycerol or ethylene glycol, is another widely adopted strategy to decrease the dephasing rate[174]. A further step it the full deuteration of the protein under investigation, which can dramatically extend the T_m, and consequently the maximum detectable distance between two spin labels[99]. Deuteration of the nitroxide spin label has been utilized in the past to mitigate intense hyperfine coupling between the electron and the spin-label methyl groups; however, the observed impact on relaxation times seems marginal in respect to protein and solvent deuteration.

5. Results

5.1. Enlarging the scenario of site directed ^{19}F labeling for NMR spectroscopy of biomolecules

In this work, we focused on developing a novel strategy to selectively incorporate ^{19}F moieties in tyrosine through post-expression covalent modification, adjusting the already known three-component Mannich type reaction. We present a successful incorporation on two distinct proteins of the commercially available parafluoroaniline tag (p-FA). The proteins that were taken under consideration are the commercially available hen egg white lysozyme (HEWL) and the immunoglobulin-binding B1 domain of group G streptococcal protein G (GB1), which was expressed and purified according to already existing protocols. The findings are supported both by ^{19}F NMR spectroscopy and ESI-MS spectrometry. Briefly, both proteins were reacted with formaldehyde and the free ^{19}F label at pH 6.5 in phosphate buffer. Through the 1D ^{19}F NMR and the ESI-MS spectra of GB1 it was possible to identify that only a single residue was involved in the conjugation reaction with a relative labelling efficiency of around 50%. This high efficiency represents a partial surprise since the Mannich reaction usually employs electron-rich anilines to better attach the carbonyl group of the formaldehyde through a nucleophilic attack. On the other hand, the HEWL showed a 1D ^{19}F NMR spectrum featuring two broad and weak peaks close to each other at around -49 ppm and another peak with higher intensity at around -50 ppm. The sharp ^{19}F NMR signal peak with higher intensity was detected even for a mixture of the protein and the correct amount of p-FA tag but without addition of formaldehyde (the needed linker between the protein and the fluorinated label) thus indicating that this signal is due to a non-covalent interaction between the fluorinated tag and the protein. Meanwhile, the broad peaks arose from the tag bound to the protein as confirmed from the ESI-MS spectra with a relative labelling efficiency of 20% on a single tyrosine residue. To increase the reaction efficiency the lysozyme was pretreated, before adding the conjugation reaction reagents, with 30% of DMSO to allow partial unfolding of the protein (confirmed by ^{1}H NMR spectra) and favor the tyrosine residue's exposure to the solvent. The ^{19}F NMR spectrum showed significative difference form the

one obtained without DMSO treatment: the peaks around -49 ppm, where greatly enhanced and a further peak appeared, hinting to the presence of a second labelling site. The ESI mass data corroborated the existence of the protein labelled on two different tyrosine residues with efficiency of nearly 50 % for the first adduct, while for the second adduct (HEWL labelled on two different resides) we had a lower efficiency.

scientific reports

Check for updates

OPEN

Enlarging the scenario of site directed ^{19}F labeling for NMR spectroscopy of biomolecules

Valentina Vitali[1,2], Francesco Torricella[1], Lara Massai[2], Luigi Messori[2] & Lucia Banci[1,2,3]✉

The possibility of using selectively incorporated ^{19}F nuclei for NMR spectroscopic studies has retrieved increasing interest in recent years. The high gyromagnetic ratio of ^{19}F and its absence in native biomolecular systems make this nucleus an interesting alternative to standard ^{1}H NMR spectroscopy. Here we show how we can attach a label, carrying a ^{19}F atom, to protein tyrosines, through the use of a specific three component Mannich-type reaction. To validate the efficacy and the specificity of the approach, we tested it on two selected systems with the aid of ESI MS measurements.

NMR spectroscopy provides atomic-level information on magnetic active nuclei. In recent years several works showed that the use of the magnetically active fluorine isotope ^{19}F represents an important alternative to the standard proton ^{1}H biomolecular NMR spectroscopy[1,2]. The 100% natural abundance, the ½ nuclear magnetic spin, and the high gyromagnetic ratio of ^{19}F represent the main features that characterize this nucleus[3]. Furthermore, the large chemical shift range of the fluorine nucleus (30 times larger than the ^{1}H nucleus) makes it a very sensitive tool for monitoring changes in its surroundings[4]. This nucleus can be easily incorporated in specific labelling sites resulting in a sensitive and powerful, yet not perturbing, probe for large and complex biological systems. Consequently, this allows the characterization in solution of biomolecules of larger molecular size than those studied through ^{1}H NMR. In recent times several applications of ^{19}F NMR for biomolecular studies such as protein–ligand interaction[5,6], protein relaxation[7], aggregation[8,9], structural and conformational changes[2], protein folding and unfolding[10,11] have been reported. Moreover, ^{19}F NMR spectroscopy is receiving increased interest even within the in-cell NMR approach, as its less crowded spectra allow to retrieve dynamical and structural information in a reliable way[8,12,13]. On this respect, Pielack et al. recently used ^{19}F NMR to describe how the intracellular environment can influence the protein dimerization equilibrium[14]. The possibility of using different ^{19}F-labelled amino acids or tags gives the chance to extend the detection of NMR signals even to crowded environments[4], as well as to exploit ^{19}F for other spectroscopic techniques, such as MAS NMR[15,16] or ENDOR[17,18]. Mainly, two kinds of approaches are used for ^{19}F nuclei incorporation in biomolecular systems: the direct expression of ^{19}F labelled proteins[19–21], or the chemical incorporation of specific molecules/labels containing fluorine nuclei[22,23]. Several works showed how the incorporation of specific fluorinated amino acids (e.g.: F-Phenylalanine, F-Tryptophan, F-Tyrosine) or unnatural amino acids[24] can be relatively easy achieved in *E.coli* cells using modified minimal media supplemented with fluorinated amino acids analogues[19]. Recently, Gotting et. al[25] presented a novel strategy to synthesize ^{13}C/^{19}F/^{2}H indoles to be used as tryptophan precursors in protein expression. On the other hand, site directed labelling (SDL) represents an attractive and potentially low-cost alternative to the direct labelling during expression, especially because the biosynthetic labelling approach, although well established, presents challenges to be overcome. In addition, the yield of the expressed protein could be significantly reduced due to the intrinsic toxicity of the fluorinated precursors[26]. Post-expression labelling can be achieved on different kinds of residues such as cysteines, lysines or tyrosines, both native or introduced with mutagenesis. Cysteine labelling is one of the most exploited approaches for the post-expression protein labelling and the use of maleimide-based tags opened up a wide variety of experimental approaches for in-cell NMR experiments[27]. A few fluorinated molecules are known to be suitable for site selective cysteine modification. Usually, these molecules are characterized by one or more trifluoromethyl groups covalently attached to a group prone to nucleophilic substitution. 3-bromo-1,1,1-trifluoroacetone (BTFA)[28–30] and analogues, 2,2,2-trifluoroethanethiol (TFET)[22,23,31,32] or 4-(perfluoro-tert-butyl)phenyliodoacetamide (PFP)[33] are among the most used molecules. Other recent and interesting developed fluorine tags were exploited to observe globular proteins directly inside human cells[34]

[1]Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy. [2]Department of Chemistry "Ugo Schiff", University of Florence, Via Della Lastruccia 3, 50019 Sesto Fiorentino, Italy. [3]Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine (CIRMMP), Florence, Italy. ✉email: banci@cerm.unifi.it

and to increase the ^{19}F chemical shift dispersion[35]. However, cysteine modification can have some drawbacks, including the fact that in a number of proteins cysteines are in the active site and/or coordinate metal cofactors.

A valuable alternative for protein labelling is tyrosine. One of the features that makes tyrosine an interesting labelling site is its average low natural abundance (just above 3%)[36], making it one of the rarest amino acids in protein sequences. Moreover, being tyrosine a partially hydrophobic residue, its location can range from being deeply buried inside the protein hydrophobic core to being surface exposed. Since the buried ones are far higher in number than the exposed ones, selecting the right reaction conditions, it could be possible to covalently label the relatively rare surface exposed tyrosines in a site-selective way[37]. Several approaches have been proposed, and recently reviewed, for the modification of tyrosine residues[37]. The most relevant ones involve the use of diazonium coupling reactions[38,39], of diazodicarboxyamides[40,41], of sulfur/fluoride exchange (SuFEx) chemistry[42] and of the Mannich-type reaction[43]. This latter reaction targets the phenolic side chain of tyrosine residues on proteins and is one of the oldest methods developed for tyrosine bioconjugation; it has been successfully used both for grafting fluorophores[43] and synthetic peptides[44] to chymotrypsinogen.

The three component Mannich-type reaction (Fig. 1) is characterized by a first step in which an imine condensation between an aldehyde and an electron-rich aromatic amine takes place. Afterwards, the phenol ring of tyrosine is deprotonated and undergoes an electrophilic aromatic substitution with the iminium ion, resulting in the formation of a novel carbon–carbon bond. This reaction was used by Francis et al. to chemically modify proteins using either small peptides or small molecules[43]. Moreover, an interesting application was reported by Belle et al. in which this reaction was used to selectively incorporate a novel spin label for EPR spectroscopy experiments[45].

In this work we report the protocol for tyrosine protein labelling using parafluoroaniline (p-FA) whose efficacy has been tested, through ESI mass spectrometry and ^{19}F NMR measurements, on two proteins of different size.

Results and discussion

p-FA tyrosine conjugation

The immunoglobulin binding domain of Streptococcal protein G (GB1) and Hen Egg White Lysozyme (HEWL) were selected as test proteins (Fig. 2), both proteins having three tyrosine residues, located in different positions of the protein structure.

Both proteins were reacted, exploiting proper reagents ratio, with formaldehyde and the free ^{19}F label. The pH at which the reaction is carried out plays a crucial role in the formation of the desired fluorinated tyrosine adducts. Indeed, operating at pH 6.5 is crucial for minimising possible side reactions on unwanted amino acid residues like tryptophans. Moreover, at this pH value the equilibrium that characterizes the reaction, could be driven mostly towards the formation of the open ring Mannich adduct. The reacted samples were then analysed

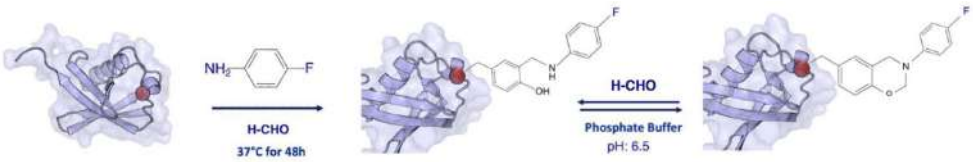


Figure 1. Reaction scheme. General representation of three component Mannich type reaction on tyrosine residue.

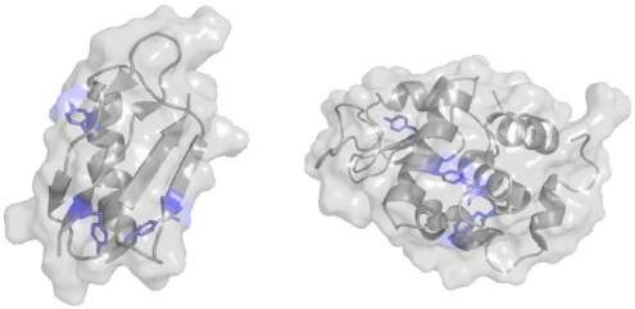


Figure 2. Protein structures. Immunoglobulin binding domain of Streptococcal protein G (GB1) (PDB: 1GB1) and Hen Egg White Lysozyme (HEWL) (PDB: 2VB1). Highlighted in blue their tyrosines.

by ^{19}F NMR to assess the presence of the fluorinated tag conjugated to the protein tyrosines, and to estimate the overall amount of fluorine nuclei conjugated onto the proteins and the number of tyrosines effectively involved in the conjugation reaction. The attachment of the fluorinated tag was further investigated by ESI–MS spectra of the intact protein before and after the coupling reaction. Mass spectrometry data were used to verify the efficiency of the conjugation reaction and the number of residues to which the tag is attached.

GB1

The 1D ^{19}F NMR spectrum of GB1 shows the presence of one well defined main peak and two smaller peaks (Fig. 3b), exhibiting different features, both in terms of shape and chemical shift, than the free fluorinated tag signal (Fig. 3a). However, it is impossible to establish the exact number of residues involved in the conjugation reaction by relying just on these NMR spectra.

It is feasible that the high intensity signals arise from one labelled tyrosine, but only the ESI–MS spectra clearly indicated (Fig. 4) that a single residue has been successfully labelled. The upfield smaller peaks observed in the ^{19}F NMR spectra can be associated with the non-covalent interactions between the protein and a small fraction of p-FA that cannot efficiently be removed during the purification steps of the reaction probably due to π-π stacking interactions between the aromatic ring of the tag and the aromatic rings of other residues.

The mass spectra data (Fig. 4b) indicate that the native unlabelled protein is still the predominant species; yet, a new peak is observed with a mass increase of 123 Dalton (50% intensity of the main peak). This peak originates from GB1 (native unlabelled protein Fig. 4a) with the p-FA tag attached to one residue which, according to the molecular weight increase, is forming the open ring adduct. By assuming an equal ionization efficiency for both species, we can directly assess the ratio of labelled to unlabelled protein which resulted to be 50:100. This high efficiency represents a partial surprise since the Mannich reaction usually employs electron-rich anilines to better attach the carbonyl group of the formaldehyde through a nucleophilic attack. The adduct was obtained through several optimization steps of the reaction conditions, such as the time of the reaction, the temperature, and the ratio of the reagents.

To identify the specific tyrosine modified by the bioconjugation reaction, we performed 1H-13C HSQC NMR spectra on both the native, unlabelled protein and the fluorinated protein. The comparison of the spectra (Supporting Information S3), suggests that the tag is attached to tyrosine 3.

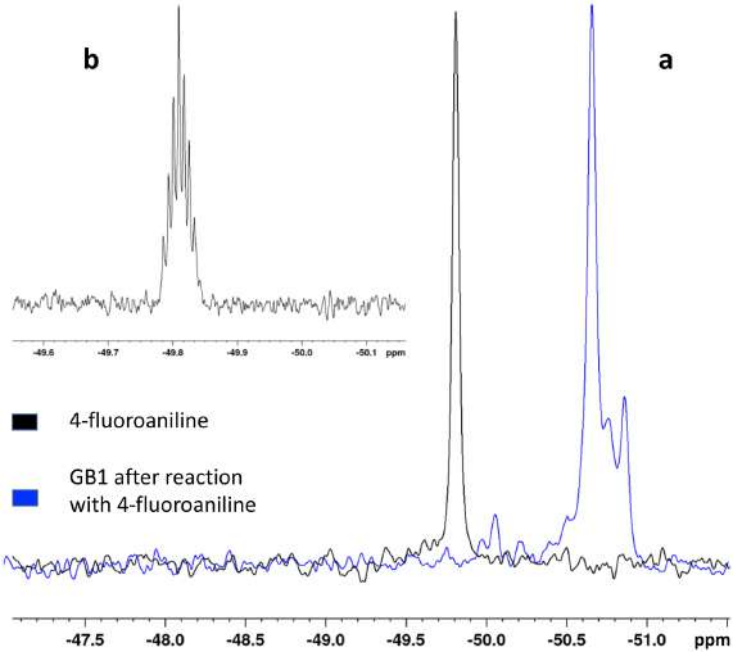


Figure 3. GB1 ^{19}F NMR spectrum. (**a**) Comparison between ^{19}F NMR spectra of p-FA (black) and GB1 after the conjugation reaction with p-FA (blue). (**b**) 4-fluoroaniline spectrum processed with a line broadening of 1Hz, showing distinct heteronuclear coupling between ^{19}F and ^{1}H nuclei.

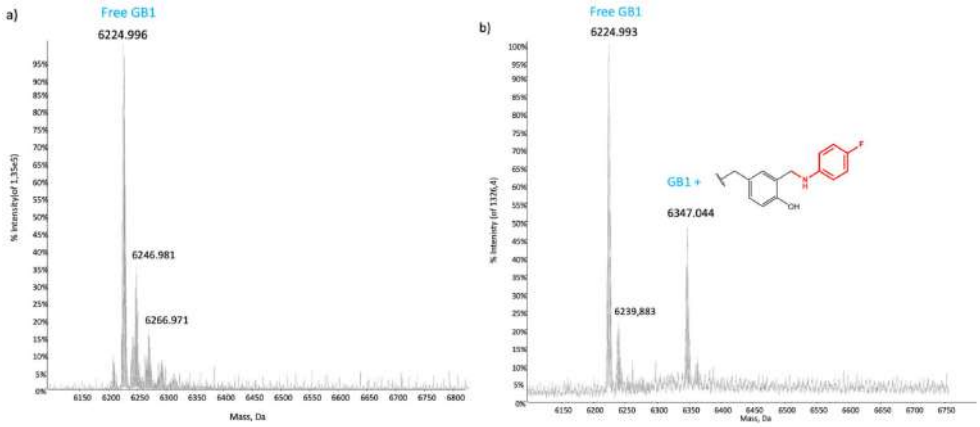


Figure 4. ESI–MS spectra of GB1. Deconvoluted ESI mass spectra of (**a**) GB1, 10^{-6} M in ammonium acetate and (**b**) GB1, 10^{-6} M, after the reaction with fluorinated tag. The peak at 6347 Da represents the GB1 open ring adduct. The bound fragment is red in the drawn structure.

HEWL

A tagged HEWL sample showed a 1D ^{19}F NMR spectrum featuring two broad, very weak peaks close to each other at around -49 ppm, and another peak with higher intensity (Fig. 5a) at around -50 ppm. At first glance, this spectrum might suggest that the latter peak originates from an effectively ^{19}F-tagged protein tyrosine and that the weaker and broader peaks are due to the low level tagging of the two other tyrosines. However, the intense and sharp ^{19}F NMR signal at -50 ppm is detected even for a mixture of the protein and the correct amount of p-FA tag but without addition of formaldehyde (the needed linker between the protein and the fluorinated label) thus indicating that this signal is due to a non-covalent interaction between the fluorinated tag and the protein, while the broad peaks could arise from the tag bound to the protein. Mass spectrometry data (Fig. 6b) confirmed that a single tyrosine among the three of lysozyme (native unlabelled protein Fig. 6a) was modified and that the two distinct broad peaks of the NMR spectrum could originate from a coexistence between the open and closed ring Mannich adduct. The existence of this equilibrium was confirmed through the comparison between simulated and experimental isotopic patterns of the sample under investigation, with peaks at + 123 Dalton for the open ring adduct and at + 137 Dalton for the closed one. Nevertheless, in this case, the reaction efficiency was significantly lower than for the GB1 protein, probably due to the reduced exposure of the tyrosine residues on the protein surface.

To further corroborate this hypothesis, HEWL was treated with 30% DMSO before adding the conjugation reaction reagents. Addition of DMSO induces a partial unfolding of the protein thus increasing the solvent exposition of the residues, including tyrosine, and leading to an increase in the reaction efficiency. The 1D ^{1}H NMR spectrum was exploited to confirm the partial unfolding of HEWL after the addition of DMSO (Supporting Information S1). Partially unfolded HEWL was then subjected to the p-fluoroaniline labelling procedure through the Mannich reaction, following the same protocol and time scheme used for the completely folded

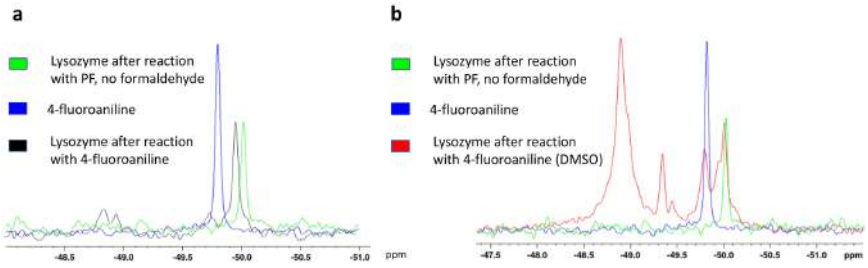


Figure 5. ^{19}F NMR spectra of lysozyme. (**a**) ^{19}F NMR spectra of p-FA (blue), lysozyme after the three component Mannich reaction (black). (**b**) Lysozyme after the three component Mannich reaction without (red) DMSO. The green spectrum represents the non-covalent interactions between the protein and the p-FA.

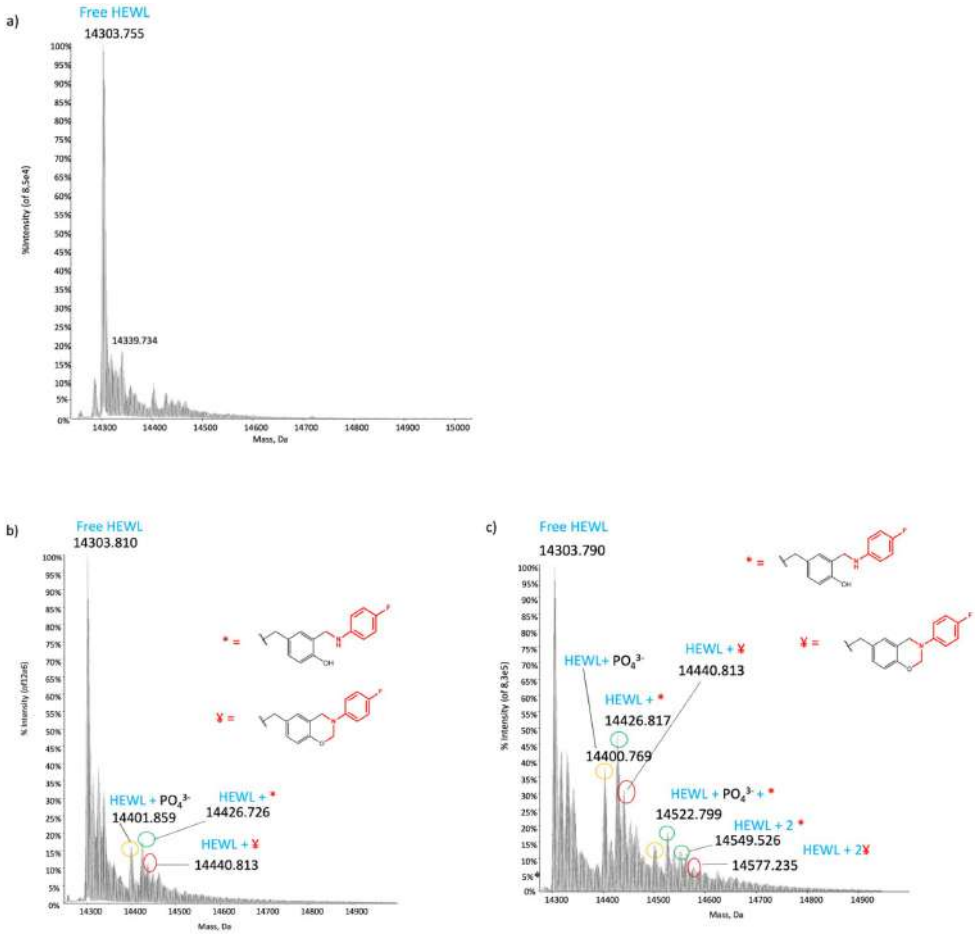


Figure 6. ESI–MS spectra of lysozyme. (**a**) Deconvoluted ESI mass spectrum of free HEWL, (**b**) deconvoluted ESI mass spectrum of HEWL after the conjugation reaction, (**c**) deconvoluted ESI mass spectrum of HEWL after the conjugation reaction with the presence of DMSO. The bound fragments are red in the drawn structures.

protein. Its ^{19}F NMR spectrum (Fig. 5b) showed significative differences with respect to that of native HEWL. The peaks between − 48.7 and − 49 ppm, already present in the sample without DMSO, are greatly enhanced; a further peak appears at − 49.3 ppm and another broader one with smaller intensity arises between − 49.4 and − 49.5 ppm. These two new peaks hint either at the possibility of labelling a second tyrosine residue or even of a third one, or to have a second tyrosine labelled both with the open and closed ring conformation. The ESI mass data corroborated the existence of the protein labelled just on two different tyrosine residues and for both of them the presence of the open and closed ring adduct, was confirmed. The ESI mass spectrum (Fig. 6c) shows a set of signals similar to those observed after tag binding to the folded protein, but with an overall increase of the labelling efficiency, with a ratio of tagged: untagged of 40:100. Moreover, the existence of a second set of signals, with a lower intensity and a higher molecular weight, confirms the presence of a second tyrosine residue labelled with the fluorinated tag.

Therefore, the use of DMSO allows gaining a larger amount of fluorinated tag attached on the tyrosine residue that is only partially labelled in the absence of DMSO.

These data confirm that only the accessible tyrosine residues can be effectively tagged and that, based on residue exposure, some selectivity in the tagging can be obtained.

Conclusions

This work presents a different application of an established mild bioconjugation reaction for NMR spectroscopy achieving the labelling of tyrosine residues with a small molecule containing the ^{19}F atom. The incorporation of a specific tag containing the ^{19}F atom offers the chance to investigate biomolecular systems in less crowded spectra compared to the ^{1}H spectroscopy. Moreover, the opportunity to insert ^{19}F atoms through a different approach than the direct overexpression of proteins with fluorinated amino acids, provides a different helpful way for situations where the direct overexpression is not applicable. We demonstrated that, using the three component Mannich-type reaction, it is possible to achieve valuable site selectivity among tyrosine residues depending on surface exposure of the tyrosine. The amphiphilicity of the phenolic side chain plays a crucial role in obtaining labelling selectivity, since most of the tyrosines are buried deep in the hydrophobic core and are not available for external modification. Therefore, both the chemical environment and the surface exposure of these residues play an important role in determining whether the residue can be labelled or not. Here, we have demonstrated that, upon adding a given amount of DMSO, HEWL can go from being labelled on one tyrosine to being labelled on two. The denaturating action of dimethyl sulfoxide exposes the residues previously inaccessible and buried inside the protein core, to the solvent and allows the conjugation reaction between the tyrosine and the fluorinated molecule. The use of a commercially available fluorinated tag such as the herein used p-fluoroaniline suggests the possible application of a wide variety of relatively cheap molecules. Moreover, the reaction yield obtained for both GB1 and HEWL with the addition of DMSO should be considered remarkable. In fact, the first step of the three component Mannich-type reaction is an imine condensation between the formaldehyde and the p-fluoroaniline. Since the imine formation starts with a nucleophilic addition of the amine to the carbonyl group, the reaction has a higher efficiency if the amine is electron rich. The fluorine atom is considered an electron withdrawing group (EWG) that decreases the electron density from the nitrogen atom and reduces the efficiency of the nucleophilic attack on the carbonyl group. In conclusion, we have demonstrated how under optimized conditions, a low-cost reaction can be exploited to perform post-expression conjugation of small fluorinated molecules to tyrosine residues. In addition, we established how the protein folding properties play a crucial role in the number of tyrosine residues than can be labelled and even in the efficiency of the reaction towards specific amino acids.

Materials and methods

GB1 T53C expression and purification

GB1 was expressed and purified according to already existing protocols[46]. Briefly, a pET-21a vector encoding for the immunoglobulin binding domain of streptococcal protein G (containing the mutation T53C) was used to transform BL21 (DE3) gold cell strain. *E. coli* cells were grown to mid-log phase at 37°C in LB medium, and then induced with 0.6 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). After induction the cells were grown for other 5 h at 20° C. The cell pellet was collected by centrifugation at 6000 rpm for 20 min and resuspend in phosphate buffer (100 mM sodium phosphate, 150 mM NaCl, pH 6.5). The suspension was heated to 80 °C, for 5 min, using a thermal bath, then cooled down on ice for 15 min and finally centrifuged at 40,000 rpm for 40 min. After filtering the supernatant, 5 mM DTT were added to the solution that was loaded onto a 16/600 Superdex 30 Increase (Cytiva) exchanging the buffer with 100 mM sodium phosphate, 150 mM NaCl, 1 mM TCEP, pH 6.5.

Hen egg white lysozyme

Hen egg white lysozyme was purchased from Sigma Aldrich.

^{19}F site directed labelling protocol

Both GB1 and HEWL were reacted with formaldehyde and p-fluoroaniline (both purchased from Sigma Aldrich) with a ratio of 1:100:30 in sodium phosphate buffer 100 mM at pH 6.5. In particular, after thawing, 100 µL of GB1 250 µM were buffer exchanged in the final phosphate reaction buffer using a PD10 desalting column. 50 µL of formaldehyde 0.25M and 30 µL (for each tyrosine) of p-fluoroaniline, were added to the protein solution. Regarding lysozyme, 3.5 mg of protein were resuspended in 1 mL of phosphate buffer for a final concentration of 250 µM. The same amount of formaldehyde and p-fluoroaniline used for GB1 were added to the HEWL solution. A second sample of HEWL was first pre-treated with 30% DMSO and then reacted with formaldehyde and p-fluoroaniline. For both proteins the reaction was incubated at 37 °C for 36h in a shaking incubator. Afterwards, the excess of p-fluoroaniline was removed through a 2.5 mL PD10 desalting column. The 3.5 ml sample volume obtained after passing through the desalting columns was concentrated exploiting a 3KDa, for GB1, and a 10KDa, for the HEWL, centricon (Merck) The volume was reduced until the protein reached 300 µM concentration. However, this method was not able to remove completely the unreacted p-FA tag, especially in the HEWL case. For this reason, a further purification of the labelled lysozyme was conducted by exploiting a gel filtration purification step. Briefly, the sample, in sodium phosphate buffer 100 mM at pH 6.5, was loaded into a size exclusion chromatography Superdex 16/60 75 pg column through a 1 mL loop. The labelled protein was collected in 1.5 mL fractions, that were concentrated to 300 µM. The 1D ^{19}F NMR spectrum conducted on this sample confirmed the complete removal of the free unreacted p-FA, (Supporting Information S2). After the purification step small aliquots of each sample were immediately taken and frozen for mass spectrometry (ESI–MS) analysis.

^{19}F NMR spectroscopy

^{19}F magnetic resonance spectra (^{19}F NMR) were recorded with a Bruker 600 MHz spectrometer with a TXI probe. Chemical shifts are reported in delta (δ) units, part per million (ppm), and were referenced to trifluoroacetic acid (TFA) as internal standard. 10% of deuterated water was added to the NMR tubes of each sample. All spectra were recorded at 298K.

MASS-spectrometry

To prepare the samples for mass analysis (ESI–MS), the 100 mM phosphate buffer pH 6.5 was exchanged with 20 mM acetate pH 6.8, with the aim of removing all kind of salts that could interfere with the experimental analysis. For the experiments with GB1 the final protein concentration was 10^{-6} M, whilst for the HEWL was 5×10^{-7} M; in all samples 0.1% v/v of formic acid was added just before sample infusion in the mass spectrometer. All measurements were carried out by direct infusion mode without the use of Liquid Chromatography (LC).

Instrumental Parameters: the ESI mass study was performed using a TripleTOF 5600 + high-resolution mass spectrometer (AB Sciex, Framingham, MA, United States), equipped with a DuoSpray® interface operating with an ESI probe. ESI mass spectra were acquired through direct sample infusion at 7 µL/min of flow rate. The general ESI source parameters optimized for the proteins analysis were as follows:

(GB1) positive polarity, ion spray voltage floating 5500V, temperature 25 °C, ion source Gas 1 (GS1) 35 L/min; ion source Gas 2 (GS2) 0; curtain gas (CUR) 20 L/min, collision energy (CE) 10 V; declustering potential (DP) 100 V, acquisition range 900–2000 m/z.

(HEWL) positive polarity, ion spray voltage floating 5500 V, temperature 25 °C, ion source Gas 1 (GS1) 40 L/min; ion source Gas 2 (GS2) 0; curtain gas (CUR) 20 L/min, collision energy (CE) 10 V; declustering potential (DP) 100 V, acquisition range 1000–2800 m/z.

For acquisition, Analyst TF software 1.7.1 (Sciex) was used, and deconvoluted spectra were obtained by using the Bio Tool Kit micro-application v.2.2 embedded in PeakViewTM software v.2.2 (Sciex).

Data availability

All data will be available from the corresponding author LB upon reasonable request.

Received: 18 September 2023; Accepted: 6 December 2023
Published online: 12 December 2023

References

1. Chen, H., Viel, S., Ziarelli, F. & Peng, L. ^{19}F NMR: A valuable tool for studying biological events. *Chem. Soc. Rev.* **42**, 7971–7982. https://doi.org/10.1039/c3cs60129c (2013).
2. Danielson, M. A. & Falke, J. J. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 163–195. https://doi.org/10.1146/annurev.bb.25.060196.001115 (1996).
3. Sharaf, N. G. & Gronenborn, A. M. (19)F-modified proteins and (19)F-containing ligands as tools in solution NMR studies of protein interactions. *Methods Enzymol.* **565**, 67–95. https://doi.org/10.1016/bs.mie.2015.05.014 (2015).
4. Gronenborn, A. M. Small, but powerful and attractive: ^{19}F in biomolecular NMR. *Structure* **30**, 6–14. https://doi.org/10.1016/j.str.2021.09.009 (2022).
5. Stadmiller, S. S., Aguilar, J. S., Waudby, C. A. & Pielak, G. J. Rapid quantification of protein-ligand binding via ^{19}F NMR lineshape analysis. *Biophys. J.* **118**, 2537–2548. https://doi.org/10.1016/j.bpj.2020.03.031 (2020).
6. Yu, L., Hajduk, P. J., Mack, J. & Olejniczak, E. T. Structural studies of Bcl-xL/ligand complexes using ^{19}F NMR. *J. Biomol. NMR* **34**, 221–227. https://doi.org/10.1007/s10858-006-0005-y (2006).
7. Lu, M., Ishima, R., Polenova, T. & Gronenborn, A. M. ^{19}F NMR relaxation studies of fluorosubstituted tryptophans. *J. Biomol. NMR* **73**, 401–409. https://doi.org/10.1007/s10858-019-00268-y (2019).
8. Zhu, W. *et al.* Visualizing proteins in mammalian cells by ^{19} F NMR spectroscopy. *Angew. Chem. Int. Ed Engl.* **61**, e202201097. https://doi.org/10.1002/anie.202201097 (2022).
9. Li, C. *et al.* ^{19}F-NMR studies of α-synuclein conformation and fibrillation. *Biochemistry* **48**, 8578–8584. https://doi.org/10.1021/bi900872p (2009).
10. Bann, J. G., Pinkner, J., Hultgren, S. J. & Frieden, C. Real-time and equilibrium ^{19}F-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD. *Proc. Natl. Acad. Sci. U S A* **99**, 709–714. https://doi.org/10.1073/pnas.022649599 (2002).
11. Li, H. & Frieden, C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and ^{19}F NMR. *Proc. Natl. Acad. Sci. U S A* **104**, 11993–11998. https://doi.org/10.1073/pnas.0705253104 (2007).
12. Li, C. *et al.* Protein ^{19}F NMR in *Escherichia coli*. *J. Am. Chem. Soc.* **132**, 321–327. https://doi.org/10.1021/ja907966n (2010).
13. Pham, L. B. T. *et al.* Direct expression of fluorinated proteins in human cells for ^{19}F in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **145**, 1389–1399. https://doi.org/10.1021/jacs.2c12086 (2023).
14. Speer, S. L. *et al.* The intracellular environment affects protein-protein interactions. *Proc. Natl. Acad. Sci. U S A* **118**. https://doi.org/10.1073/pnas.2019918118 (2021).
15. Wang, M. *et al.* Fast magic-angle spinning ^{19} F NMR spectroscopy of HIV-1 capsid protein assemblies. *Angew. Chem. Int. Ed Engl.* **57**, 16375–16379. https://doi.org/10.1002/anie.201809060 (2018).
16. Porat-Dahlerbruch, G., Struppe, J., Quinn, C. M., Gronenborn, A. M. & Polenova, T. ^{19}F fast MAS (60–111 kHz) dipolar and scalar based correlation spectroscopy of organic molecules and pharmaceutical formulations. *Solid State Nucl. Magn. Reson.* **122**, 101831. https://doi.org/10.1016/j.ssnmr.2022.101831 (2022).
17. Asanbaeva, N. B. *et al.* Application of W-band ^{19}F electron nuclear double resonance (ENDOR) spectroscopy to distance measurement using a trityl spin probe and a fluorine label. *Phys. Chem. Chem. Phys.* **24**, 5982–6001. https://doi.org/10.1039/d1cp05445g (2022).
18. Kehl, A. *et al.* Resolution of chemical shift anisotropy in ^{19}F ENDOR spectroscopy at 263 GHz/9.4 T. *J. Magn. Reson.* **333**, 107091. https://doi.org/10.1016/j.jmr.2021.107091 (2021).
19. Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of ^{19}F-Tryptophan for proteinNMR spectroscopy. *Chem. Commun. (Camb)* **48**, 10681–10683. https://doi.org/10.1039/c2cc35347d (2012).
20. Campos-Olivas, R., Aziz, R., Helms, G. L., Evans, J. N. & Gronenborn, A. M. Placement of ^{19}F into the center of GB1: Effects on structure and stability. *FEBS Lett.* **517**, 55–60. https://doi.org/10.1016/s0014-5793(02)02577-2 (2002).
21. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The preparation of 19F-labeled proteins for NMR studies. *Methods Enzymol.* **380**, 400–415. https://doi.org/10.1016/S0076-6879(04)80018-1 (2004).
22. Susac, L., Eddy, M. T., Didenko, T., Stevens, R. C. & Wuthrich, K. A_{2A} adenosine receptor functional states characterized by ^{19}F-NMR. *Proc. Natl. Acad. Sci. U S A* **115**, 12733–12738. https://doi.org/10.1073/pnas.1813649115 (2018).
23. Klein-Seetharaman, J., Getmanova, E. V., Loewen, M. C., Reeves, P. J. & Khorana, H. G. NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: Applicability of solution (19)F NMR. *Proc. Natl. Acad. Sci. U S A* **96**, 13744–13749. https://doi.org/10.1073/pnas.96.24.13744 (1999).

24. Oyala, P. H. *et al.* Biophysical characterization of fluorotyrosine probes site-specifically incorporated into enzymes: *E. coli* ribonucleotide reductase as an example. *J. Am. Chem. Soc.* **138**, 7951–7964. https://doi.org/10.1021/jacs.6b03605 (2016).
25. Maleckis, A., Herath, I. D. & Otting, G. Synthesis of $^{13}C/^{19}F/^{2}H$ labeled indoles for use as tryptophan precursors for protein NMR spectroscopy. *Org. Biomol. Chem.* **19**, 5133–5147. https://doi.org/10.1039/d1ob00611h (2021).
26. Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of 19F NMR to studies of protein structure and dynamics. *Prog. Nucl. Magn. Reson. Spectrosc.* **62**, 1–33. https://doi.org/10.1016/j.pnmrs.2011.06.003 (2012).
27. Karthikeyan, G. *et al.* A bioresistant nitroxide spin label for in-cell EPR spectroscopy: In vitro and in oocytes protein structural dynamics studies. *Angew. Chem. Int. Ed Engl.* **57**, 1366–1370. https://doi.org/10.1002/anie.201710184 (2018).
28. Luchette, P. A., Prosser, R. S. & Sanders, C. R. Oxygen as a paramagnetic probe of membrane protein structure by cysteine mutagenesis and ^{19}F NMR spectroscopy. *J. Am. Chem. Soc.* **124**, 1778–1781. https://doi.org/10.1021/ja016748e (2002).
29. Thomas, M. R. & Boxer, S. G. 19F NMR of trifluoroacetyl-labeled cysteine mutants of myoglobin: Structural probes of nitric oxide bound to the H93G cavity mutant. *Biochemistry* **40**, 8588–8596. https://doi.org/10.1021/bi0101087 (2001).
30. Hellmich, U. A., Pfleger, N. & Glaubitz, C. F-MAS NMR on proteorhodopsin: Enhanced protocol for site-specific labeling for general application to membrane proteins. *Photochem. Photobiol.* **85**, 535–539. https://doi.org/10.1111/j.1751-1097.2008.00498.x (2009).
31. Loewen, M. C. *et al.* Solution ^{19}F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. *Proc. Natl. Acad. Sci. U S A* **98**, 4888–4892. https://doi.org/10.1073/pnas.051633098 (2001).
32. Bondarenko, V. *et al.* ^{19}F paramagnetic relaxation-based NMR for quaternary structural restraints of ion channels. *ACS Chem. Biol.* **14**, 2160–2165. https://doi.org/10.1021/acschembio.9b00692 (2019).
33. Kalbitzer, H. R. *et al.* A new high sensitivity 19F probe for labeling cysteine groups of proteins. *NMR Biomed.* **5**, 347–350. https://doi.org/10.1002/nbm.1940050605 (1992).
34. Chai, Z. *et al.* Visualizing proteins in human cells at near-physiological concentrations with sensitive ^{19}F NMR chemical tags. *Angew. Chem. Int. Ed Engl.* **62**, e202300318. https://doi.org/10.1002/anie.202300318 (2023).
35. Huang, Y. *et al.* Environmentally ultrasensitive fluorine probe to resolve protein conformational ensembles by ^{19}F NMR and Cryo-EM. *J. Am. Chem. Soc.* **145**, 8583–8592. https://doi.org/10.1021/jacs.3c01003 (2023).
36. Ischiropoulos, H. Biological tyrosine nitration: A pathophysiological function of nitric oxide and reactive oxygen species. *Arch. Biochem. Biophys.* **356**, 1–11. https://doi.org/10.1006/abbi.1998.0755 (1998).
37. Szijj, P. A., Kostadinova, K. A., Spears, R. J. & Chudasama, V. Tyrosine bioconjugation—an emergent alternative. *Org. Biomol. Chem.* **18**, 9018–9028. https://doi.org/10.1039/d0ob01912g (2020).
38. Hooker, J. M., Kovacs, E. W. & Francis, M. B. Interior surface modification of bacteriophage MS2. *J. Am. Chem. Soc.* **126**, 3718–3719. https://doi.org/10.1021/ja031790q (2004).
39. Schlick, T. L., Ding, Z., Kovacs, E. W. & Francis, M. B. Dual-surface modification of the tobacco mosaic virus. *J. Am. Chem. Soc.* **127**, 3718–3723. https://doi.org/10.1021/ja046239n (2005).
40. Ban, H. *et al.* Facile and stabile linkages through tyrosine: Bioconjugation strategies with the tyrosine-click reaction. *Bioconjug. Chem.* **24**, 520–532. https://doi.org/10.1021/bc300665t (2013).
41. Bruins, J. J. *et al.* Inducible, site-specific protein labeling by tyrosine oxidation-strain-promoted (4 + 2) cycloaddition. *Bioconjug. Chem.* **28**, 1189–1193. https://doi.org/10.1021/acs.bioconjchem.7b00046 (2017).
42. Choi, E. J., Jung, D., Kim, J. S., Lee, Y. & Kim, B. M. Chemoselective tyrosine bioconjugation through sulfate click reaction. *Chemistry* **24**, 10948–10952. https://doi.org/10.1002/chem.201802380 (2018).
43. Joshi, N. S., Whitaker, L. R. & Francis, M. B. A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J. Am. Chem. Soc.* **126**, 15942–15943. https://doi.org/10.1021/ja0439017 (2004).
44. Romanini, D. W. & Francis, M. B. Attachment of peptide building blocks to proteins through tyrosine bioconjugation. *Bioconjug. Chem.* **19**, 153–157. https://doi.org/10.1021/bc700231v (2008).
45. Lorenzi, M. *et al.* Tyrosine-targeted spin labeling and EPR spectroscopy: An alternative strategy for studying structural transitions in proteins. *Angew. Chem. Int. Ed Engl.* **50**, 9108–9111. https://doi.org/10.1002/anie.201102539 (2011).
46. Schmidt, H. L. *et al.* Crystal polymorphism of protein GB1 examined by solid-state NMR spectroscopy and X-ray diffraction. *J. Phys. Chem. B* **111**, 14362–14369. https://doi.org/10.1021/jp075531p (2007).

Acknowledgements
This work was supported by iNEXT-Discovery, Grant No.871037, funded by the Horizon 2020 research and innovation programme of the European Commission, by Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italian Instruct Centre, by Ministero dell'Università e della Ricerca PRIN Grant No. 20177XJCHX. LM and LM thank A. Geri for the help in the mass spectra editing.

Author contributions
V.V.: conceptualization, investigation, data curation, formal analysis, writing-original draft, writing—review and editing. F.T.: conceptualization, investigation, data curation, formal analysis, writing—original draft, writing—review and editing. L.Ma.: investigation, data curation, writing—review and editing. L.M.: writing—review and editing. L.B.: conceptualization, funding acquisition, supervision, writing—review and editing.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-49247-2.

Correspondence and requests for materials should be addressed to L.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supporting Info

Enlarging the scenario of site directed ^{19}F labeling for NMR spectroscopy of biomolecules

Valentina Vitali[1,2], Francesco Torricella[1], Lara Massai[2], Luigi Messori[2], and Lucia Banci*[1,2,3]

[1]Magnetic Resonance Center (CERM), University of Florence, via Luigi Sacconi 6, Sesto Fiorentino 50019, Italy; [2]Department of Chemistry "Ugo Schiff", University of Florence, via della Lastruccia 3, Sesto Fiorentino 50019, Italy; [3]Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine (CIRMMP), Florence, Italy

*Corresponding author
Lucia Banci: banci@cerm.unifi.it

Table of Contents

1. GB1 protein sequence

Full amino-acid protein sequence for Immunoglobulin Binding Domain of Protein G, with T53C mutation, where the three tyrosine residues highlighted in red.

```
1          10          20          30          40          50
MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFCVTE
```

2. ^{1}H 1D NMR spectra of Hen Egg White Lysozyme (HEWL)

Here reported (**Figure S1**) the two 1D ^{1}H NMR spectra of HEWL before and after the treatment with 30% DMSO. As expected [1], the overall chemical shift dispersion in the ^{1}H NMR spectrum of lysozyme in 30% DMSO, is reduced compared to the native state with broadening of several signals. Moreover, in the region between 5 and 6 ppm we observe a disappearance of peaks which is diagnostic for a partial unfolding state of the protein. We recorder another 1D 1H NMR spectrum after the removal of DMSO and as expected the original folding of the protein was completely restored.

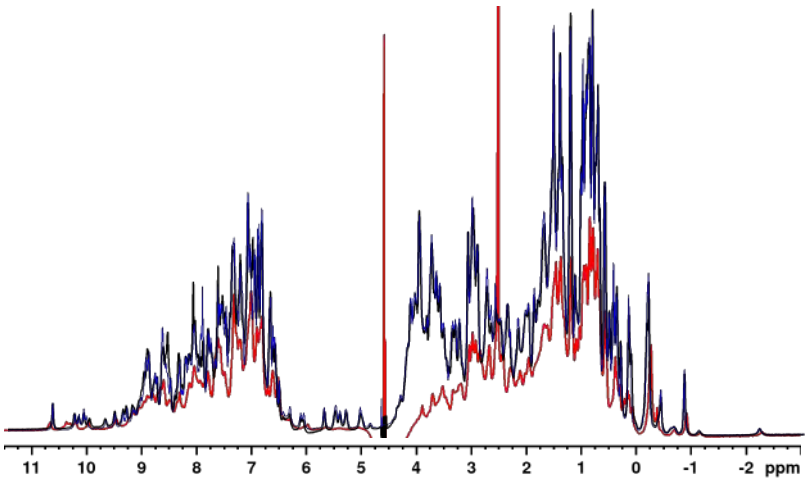


Figure S1 *Lysozyme 1D ^{1}H NMR spectra, before (red) and after (black) 30% DMSO treatment. In blue the spectrum of the HEWL after removing the DMSO.*

All spectra were recorded at 700 MHz, and 25°C with 10% of deuterated water D_2O.

3. Purification after labelling reaction

To enhance the purification step instead of a single PD10 the protein was loaded to a gel filtration column Superdex 16/600 75 pg. The figure below shows the acquired 1D ^{19}F spectrum of the HEWL ^{19}F labelled sample after this additional purification step.

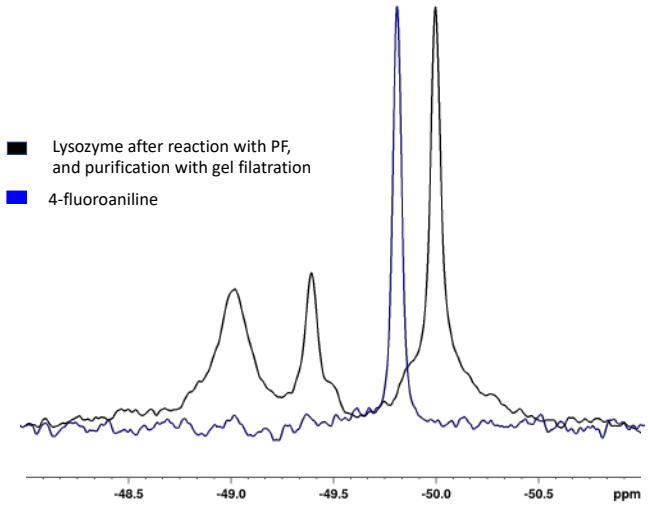


Figure S2 *1D ^{19}F NMR spectrum of HEWL after purification with a size exclusion chromatography column (in black), while in blue the 1D ^{19}F NMR spectrum of the free p-FA label.*

4. 1H-13CHSQC NMR spectra

To determine which of the three tyrosine residues has been labelled in GB1, we acquired 1H-13C HSQC NMR spectra on both the native protein (in red) and the labeled fluorinated protein (in blue) (**Figure S3)**. Our focus was on the aromatic region of the spectrum: as the tag is attached to the epsilon carbon of tyrosine, its signal is expected to be affected. By comparing the normalized intensities of the peaks corresponding to the three tyrosines, we observed a relatively large intensity variation only in the peak associated with the epsilon carbon of tyrosine 3 (**Figure S4).**

The spectra assignment is taken from BMRB code n. 25909

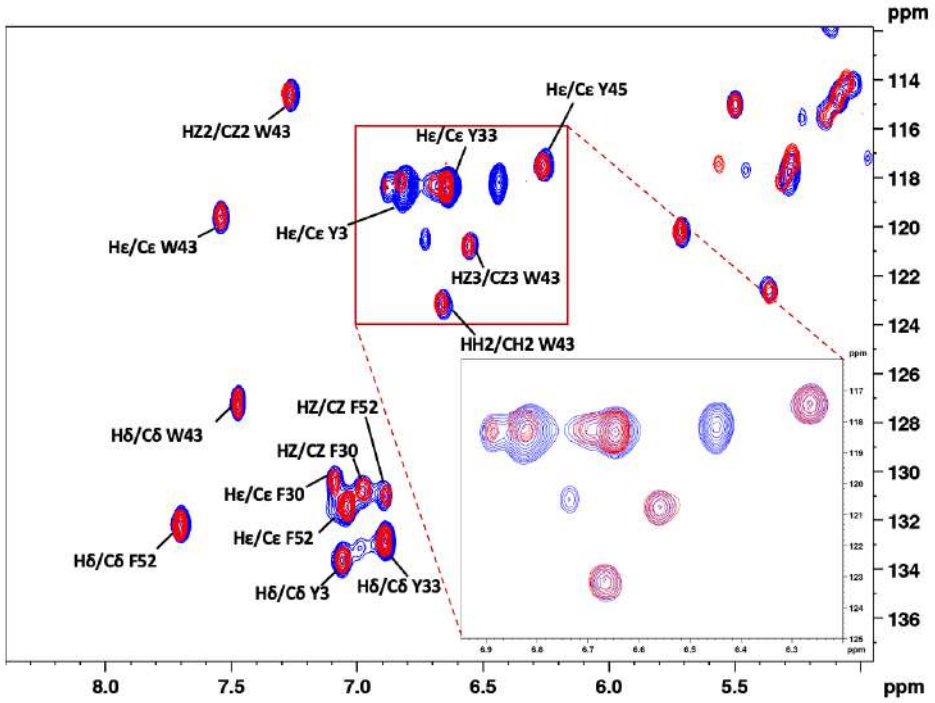


Figure S3 *1H-13C HSQC NMR spectra of the GB1 protein. In red the spectrum of the native protein, in blue the spectrum of the protein after the labelling reaction with the p-FA. The peak belonging to the carbon epsilon of the tyrosine 3. The peaks at 118.20, 6.45 ppm and the one at 120.5 and 6.7 ppm are only present in the blue spectrum and represent the aromatic ring signals of the p-FA tag. All spectra were acquired under the same conditions at 500MHz and 25°C with 10% of deuterated water D_2O for every sample.*

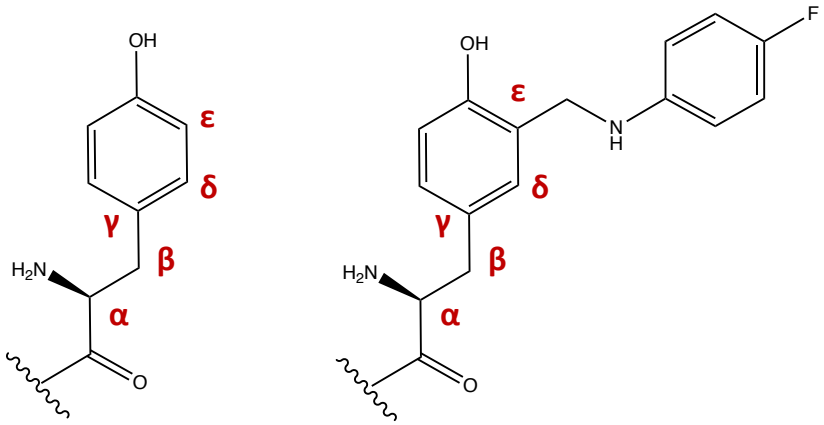


Figure S4 *Schematic representation of the chemical structure of tyrosine residues with and without the fluorinated label. In red the nomenclature used for defining the carbon and hydrogen atoms.*

5. Reference

1 Bhattacharjya, S. & Balaram, P. Effects of organic solvents on protein structures: observation of a structured helical core in hen egg-white lysozyme in aqueous dimethylsulfoxide. *Proteins* **29**, 492-507 (1997). https://doi.org:https://doi.org/10.1002/(SICI)1097-0134(199712)29:4%3C492::AID-PROT9%3E3.0.CO;2-A

4

5.2. Spectroscopically Orthogonal Labelling to Disentangl Site-Specific Nitroxide Label Distributions

In this work, the characteristic of four different nitroxide spin labels were extensively tested exploiting two constructs of the small model system GB1 (I6C/K28H/Q32H and I6H/N8H/K28C), respectively with a single cysteine residue on the β-sheet and a double histidine motif on the α-helix and vice versa. The four tested labels were MTSL, MPSL, IPSL and IDSL. The double histidine motif, being highly rigid, coupled with Cu(II) chelate complexes provides narrower distance distribution, with respect to the nitroxide-nitroxide case. Here we employed the combination of nitroxides and copper chelator ligand (CuNTA). Their distances were extracted with both ct and vt RIDME pulse sequences. The width and the shape of the corresponding distance distributions helped in disentangling the individual behaviour of nitroxide labels attached either on the α-helix or the β-sheet. Afterwards, two *in silico* labelling strategies (MMM and MtsslWizard) were employed to check whether modelling approaches could be helpful in predicting distance distributions by properly reproducing the experimental ones, and to address possible discrepancies between the two methods. We tested these *in silico* methods both for an experimental x-ray crystallography structure (4wh4) and for three structures predicted respectively by AlphaFold2, OmegaFold and ESMFold. Briefly, the experimental distance distributions extracted from the ct and vt RIDME experiments do not display significant discrepancies, as expected for short distances and relatively narrow distance distributions (as in this case). However, the sensitivity per echo and per unit of time confirmed that the vtRIDME gave generally higher sensitivity values than the ctRIDME. Moreover, the distance distributions of the GB1 with the nitroxide on the β-sheet site were characterized by monomodal and relatively narrow distributions, with hints of a shoulder for the MPSL and IPSL labels. On the other hand, all nitroxide labels on the α-helix (except IDSL) displayed a bimodal trend with two distinct and large amplitude peaks arising from the presence of two different sub-ensembles of conformers. In fact, nitroxides with long enough linkers and discrete flexibility can be positioned pointing away from the protein back-bone at the maximum distance possible, but also bent towards the protein surface, resulting in shorter distances. Therefore, IDSL, having

the shortest and least flexible linker is the only label which retains unimodal and relatively narrow distance distribution independently of the side of the GB1 on which it is attached. The comparison between the in *silico* and the experimental distance distributions revealed an intriguing difference between MMM and MtsslWizard. While the former consistently underestimated the distances between the two labels, with respect to the experimental data, the latter tended to overestimate the same measurements. both approaches seemed more accurate in the prediction of the narrow and unimodal distributions of the nitroxide labelling on the β-sheet, with the only exception being MMM at cryogenic temperature settings, that simulated hints of bimodality for MPSL. On the other hand, the predicted distributions for nitroxide labelling on the α-helix were characterized by an increased broadness that partially covered the two distinct peaks of the experimental data, although neither MMM nor MtsslWizard seemed to fully and reliably predict the bimodality of the experimental data. To better evaluate the performance of the *in silico* approaches we built different correlation plots by extracting the mean and width of every distance distribution from the experimental and the *in silico* labelling data. Altogether, neither MMM nor MtsslWizard clearly outperformed the other. For the I6C/K28H/Q32H construct MTSL was predicted best by MMM at ambient temperature, while all other nitroxides were best predicted by MtsslWizard Loose. On the other hand, the interpretation of the data for the I6H/N8H/K28C was not as straightforward. Additional correlation plots were created to evaluate how the prediction tools affected the *in silico* labelling performances. In general, predicting the distance distributions between the CuNTA and the nitroxide on the β-sheet site seemed to be mostly unaffected by the choice of the structure prediction method (AlphaFold2, OmegaFold, and ESMFold), as expected for a small, well-known globular protein, like GB1. In contrast, the *in silico* spin labelling approach clearly had a greater impact on predicted distance distributions. However, the situation was found to be more complex for the I6H/N8H/K28C construct, where the choice of the model prediction influenced the outcome of the distance distribution more strongly, although always with a lower impact than the influence derived from the labelling choice.

Applied Magnetic Resonance
https://doi.org/10.1007/s00723-023-01611-1

Applied
Magnetic Resonance

ORIGINAL PAPER

Spectroscopically Orthogonal Labelling to Disentangle Site-Specific Nitroxide Label Distributions

Valentina Vitali[1,2,3] · **Katrin Ackermann**[1] · **Gregor Hagelueken**[4] · **Bela E. Bode**[1]

Received: 30 June 2023 / Revised: 23 August 2023 / Accepted: 25 August 2023

Abstract
Biomolecular applications of pulse dipolar electron paramagnetic resonance spectroscopy (PDS) are becoming increasingly valuable in structural biology. Site-directed spin labelling of proteins is routinely performed using nitroxides, with paramagnetic metal ions and other organic radicals gaining popularity as alternative spin centres. Spectroscopically orthogonal spin labelling using different types of labels potentially increases the information content available from a single sample. When analysing experimental distance distributions between two nitroxide spin labels, the site-specific rotamer information has been projected into the distance and is not readily available, and the contributions of individual labelling sites to the width of the distance distribution are not obvious from the PDS data. Here, we exploit the exquisite precision of labelling double-histidine (dHis) motifs with Cu^{II} chelate complexes. The contribution of this label to the distance distribution widths in model protein GB1 has been shown to be negligible. By combining a dHis Cu^{II} labelling site with cysteine-specific nitroxide labelling, we gather insights on the label rotamers at two distinct sites, comparing their contributions to distance distributions based on different in silico modelling approaches and structural models. From this study, it seems advisable to consider discrepancies between different in silico modelling approaches when selecting labelling sites for PDS studies.

1 Introduction

Understanding topologies and conformational transitions of biomacromolecules and their role in health and disease has spurred a growing scientific interest in pulse dipolar electron paramagnetic resonance spectroscopy (PDS) as a reliable source for structural insights into proteins [1–3] and nucleic acids [4–6]. PDS is often combined with other biophysical methods, including X-ray crystallography, cryo-electron microscopy, or Nuclear Magnetic Resonance (NMR) spectroscopy. PDS derives

Extended author information available on the last page of the article

Published online: 24 September 2023

its strength, efficacy, and versatility from depending solely on the presence of paramagnetic centres, while being unaffected by the complexity or the size of the diamagnetic part of the system under investigation.

These paramagnetic probes are typically attached to biomacromolecules through Site-Directed Spin Labelling (SDSL), an approach pioneered during the late 1980s by Hubbell and co-workers [7–9], based on the introduction of small spin-bearing molecules at specific and strategically placed sites within the native diamagnetic system. PDS is employed to retrieve the magnetic dipole–dipole interaction between two or more spin labels and has become widely and successfully applied for retrieving interatomic distances in the nanometre range (about 2–10 nm and even beyond) [10, 11]. Obtaining information about the distance distributions between pairs of spin labels is crucial for understanding the structural organization [12, 13], dynamics, and conformational changes [14–18] of a vast array of biomolecules.

Depending on the specific case studied, a growing number of spin labels with different properties have been developed. Paramagnetic labels commonly employed for PDS include nitroxides, transition metal ions (such as copper [19, 20] and manganese [21, 22] ions), and lanthanide ions (such as gadolinium) [23, 24], or native metal clusters [25]. Nevertheless, despite this huge variety of types of spin labels, nitroxide radicals are by far the most commonly adopted and are typically attached through covalent bonds to cysteine residues introduced via SDSL. The unique chemistry of the thiol group of this amino acid and its general low abundance in protein sequences results in a high selectivity of the conjugation reaction. Among the great pool of available nitroxide labels, four have recently been tested in PDS on the model system immunoglobulin-binding B1 domain of group G streptococcal protein G (GB1) with two cysteine mutations (one on the α-helix and the other in the β-sheet) [26]. The four tested labels were: (1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanethiosulfonate (MTSL or MTSSL) [27, 28], 3-Maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (MPSL) [29–31], 3-(2-Iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (IPSL) [32–34], and bis-(2,2,5,5-Tetramethyl-3-imidazoline-1-oxyl-4-yl)disulfide (IDSL) [35–38] (Fig. 1a). Each one possesses distinct reactivity and chemical properties that enable unique applications in probing various aspects of biomolecular systems.

Notably, MTSL is the most robust and widespread used nitroxide label. It can be grafted onto the site of choice with high specificity. MPSL is characterized by a greater conformational flexibility due to the increased length of its linker. It still reacts easily with the target protein and is less prone than MTSL to be reductively cleaved due to the absence of the disulfide bond. On the other hand, IPSL tends to be less reactive and requires higher excess for quantitative labelling [26]. The biradical IDSL has the shortest linker that is least flexible, it is also the least stably attached and requires the largest excess of the four nitroxides in labelling reactions. Almost 30-fold excess of label, with respect to the free cysteines, is required due to the equilibrium constant of the disulfide exchange, between the free label disulfide bond and the disulfide bond between the cysteine and the label, being close to one [26]. In the previous study [26], some of the extracted distance distributions between the two nitroxides were broad or even bimodal; however, in samples doubly labelled with nitroxides, it is not straightforward to

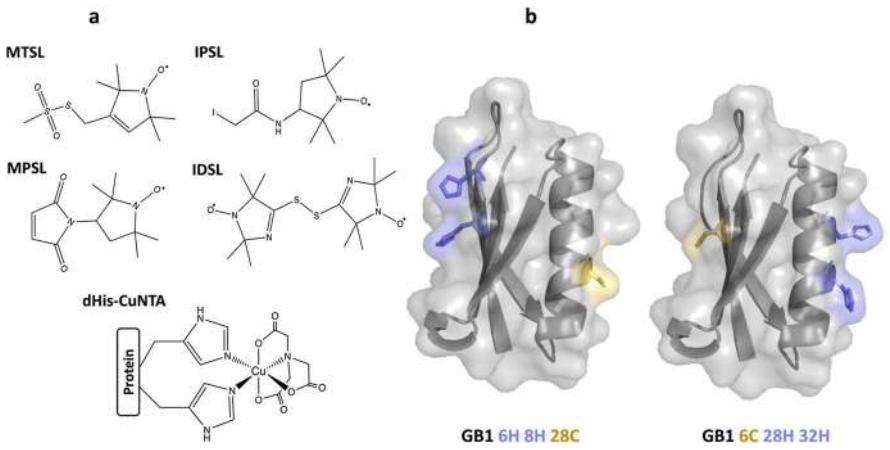


Fig. 1 **a** Structures of the four free nitroxide labels used for conjugation of the cysteine residues, respectively, in position 28 (α-helix) and 6 (β-sheet) of the GB1 protein and the CuNTA label directly coordinating the dHis motif (dHis-CuNTA). **b** The two GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C employed in this study

assign the contributions of the individual labelling site to the distribution width and shape. However, the bipedal spin labels formed by metal chelate complexes of copper(II) such as copper nitrilotriacetate (CuNTA) [39] or copper iminodiacetic acid (CuIDA) [40] bound to double-histidine (dHis) motifs engineered into specific sites [41] display a much narrower distance distribution in GB1 [20]. As observed for the tetra-histidine GB1 construct (I6H/N8H/K28H/Q32H), the intrinsic rigidity and the lower conformational flexibility of this spin pair ensures narrower and more precise distance distributions between the Cu^{II}–Cu^{II} pair with respect to the nitroxide–nitroxide distances, making this motif ideal for an independent investigation of the two nitroxides' behaviours. Although this labelling alternative is not based on covalent bonding, its robustness against competitor ligands has been tested extensively [42].

Therefore, to better understand the influence, in this case, of the nitroxide labelled secondary structural elements (α-helix *vs.* β-sheet) and their respective behaviours in PDS experiments, we have investigated each site of GB1 independently. Here, we have employed orthogonal labelling, individually combining the aforementioned nitroxides with a dHis motif [41] with bipedal coordination to CuNTA [20, 39] alternately at either labelling site, i.e., in the α-helix or the β-sheet, of GB1 (Fig. 1b).

To reliably translate the spectroscopic information on the labelled samples to the actual protein structure of interest, different approaches to explicitly model the labels onto the proteins have been developed. Modelling the spin label conformers through in silico approaches can, e.g., be useful to better understand the reasons behind the differences in widths and shapes of the distributions of the systems under investigation and to reliably translate distance distribution into conformers or conformer ensembles of proteins. The corresponding computational tools provide means to study spin-labelled biomolecules by modelling the distance distributions

between paramagnetic moieties and are often employed to identify suitable labelling sites by performing a systematic scanning of the protein residues.

Here, we have explored the two different in silico labelling approaches MMM [43, 44] and MtsslWizard [45], built on different theoretical concepts. While MtsslWizard is based on the excluded volume by steric clashes between label rotamers and the protein, MMM relies on the Lennard Jones potential energy and intrinsic energies of spin label rotamers. The new modelling package chiLife implements these two approaches in a single software tool [46] and has very recently incorporated the use of bipedal labels [47]. Beyond giving information on the distributions between pairs of nitroxide labels, MMM has recently introduced the possibility to simulate the labelling of a dHis motif with coordination complexes such as CuNTA and to retrieve the corresponding distance distributions [44]. Here, we introduce and test a new implementation for bipedal labels into MtsslWizard. The original version of the software superimposes a model of a spin label onto an amino acid (or nucleobase) residue and generates an ensemble of non-clashing rotamers [45]. In essence, this procedure generates an estimate of the accessible volume of the spin label. While this is straightforward for monopedal labels, such as MTSL, bipedal labels are harder to model. Here, the chi angles cannot be randomly chosen, because the other end of the label must coincide with the main chain atoms of the second labelling site.

We have designed a genetic search algorithm to solve this problem. We found that this algorithm quickly converges to a set of rotamers that connect the two labelling sites. As for the published version of MtsslWizard, parameters such as "tight" or "loose" van-der-Waals cutoff can be selected. In principle, the algorithm should work for any bipedal label. Currently, the parameters for the Cu-based labels in Fig. 1 have been implemented (see supplementary.pse file). The code is freely available on GitHub [48] and as a colab notebook [49]. It will be included in a future update of the MtsslSuite-Server (http://www.mtsslsuite.isb.ukbonn.de/).

When studying structures and conformational flexibilities of proteins by SDSL and PDS, high-quality high-resolution structures or structural models are crucially important. While this is commonly based on experimental structures, recent extraordinary progress in deep learning methods has sparked the development of powerful tools for protein structure prediction. Together with in silico labelling techniques, this can greatly assist in the selection of the labelling sites of a protein for PDS experiments. Herein, we have explored three different structure prediction tools: AlphaFold2, which is based on the predictions of protein folding using network-based models and multiple sequence alignments [50], OmegaFold, which leverages deep learning to predict high-resolution protein structure from a single primary sequence without depending on sequence alignment [51], and ESMFold, which predicts protein folding using large-scale language model on a single primary sequence and demonstrated great runtime efficiency [52]. The outcomes of these prediction tools were compared to the ones obtained using one of the GB1 crystallographic structures (PDB: 4wh4) [20].

From the experimental point of view, the significant spectral separation between dHis-CuNTA and nitroxides makes Relaxation-Induced Dipolar Modulation Enhancement (RIDME) [53–55] the natural choice for investigating the distance

distributions between these orthogonal labels. This is owed to the fact that the spectrum of copper(II) as a paramagnetic metal centre is much broader than that of a nitroxide radical. Both spectra are also fully separated (i.e., non-overlapping) at Q-band EPR frequencies. Thus, double-resonance techniques that require selective excitation of both spin centres are only feasible with very large bandwidth not available in most EPR resonators. Single frequency techniques that require the excitation of both spins are even less feasible as the excitation would require covering both spectra at the same time. In RIDME, however, the fast-relaxing spin is inverted statistically by longitudinal relaxation introducing the dipolar interaction. Thus, the microwave pulses only need to excite a sufficient fraction of the nitroxide radical to detect the echo which is readily achieved in most setups. By analyzing the modulation with the dipolar frequency, the inter-spin distance can be recovered. RIDME has the further advantage that orientation-dependent excitation of the broad copper(II) spectrum can be avoided as all orientations will contribute, simplifying analysis and interpretation [56]. Here, we compare the 5-pulse variable-time (vt) RIDME, recently introduced by our group [57], with the standard 5-pulse constant-time (ct) RIDME. Testing the vtRIDME sequence further will allow evaluating its performance compared to the constant-time sequence, to gain a deeper understanding of its potential and limitations.

2 Experimental Section

2.1 Expression, Purification, and Spin Labelling

The model protein used for this study is the small and rigid immunoglobulin-binding B1 domain of group G streptococcal protein G (GB1). Both GB1 constructs investigated here have a cysteine for the nitroxide labelling and a dHis motif for the chelator agent CuNTA. The I6C/K28H/Q32H construct has the cysteine residue in the β-sheet site, while the I6H/N8H/K28C construct has it in the α-helix. Both constructs were expressed and purified as previously described [19]. The nitroxide spin labels used in this study are MTSL [(1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate; Santa Cruz Biotechnology], MPSL (3-Maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxy; Santa Cruz Biotechnology), IPSL (3-(2-Iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy; Sigma-Aldrich), and IDSL (bis-(2,2,5,5-Tetramethyl-3-imidazoline-1-oxyl-4-yl)disulfide; Noxygen). These nitroxides have already been employed in our previous study where we have developed the respective labelling protocols [26], which were followed in the current study apart from two adjustments regarding the molar ratio of label to cysteine. Here, we have used 15:1 label:cysteine for IPSL and 25:1 label:cysteine for IDSL, with the aim of maximizing their labelling efficiency, while keeping ratios for MPSL and MTSL the same as used previously at 3:1 label:cysteine [26]. The labelling reaction was carried out in phosphate buffer (42.4 mM Na_2HPO_4, 7.6 mM KH_2PO_4, 150 mM NaCl, and pH 7.4) as described [26].

Successful spin labelling was confirmed via electrospray ionisation (ESI) mass spectrometry using the in-house mass spectrometry facility. Unlabelled (control) and

labelled I6H/N8H/K28C and I6C/K28H/Q32H GB1 samples were diluted to 1 µM in 1% formic acid (FA). 30 µL (30 pmol) per sample were injected onto the liquid chromatography (LC) system (Waters Xevo G2 TOF MS with Acquity HPLC) using a MassPrep cartridge column (Waters), applying a 5 min gradient from 95% water, 5% acetonitrile to 5% water, and 95% acetonitrile (eluents supplemented with 1% FA). Data were collected in positive mode from 500 to 2500 m/z, and charged ion series deconvolution to 0.1 Da resolution was performed using the MaxEnt I algorithm utilising a peak width at half height of 0.4 m/z. Results are shown in the supplementary information (SI) (Fig. S1, Table S1).

Labelling efficiency was confirmed with room-temperature continuous wave (CW) EPR. CW EPR experiments were performed using a Bruker EMX 10/12 spectrometer equipped with an ELEXSYS Super Hi-Q resonator at an operating frequency of ~ 9.9 GHz (X-band) with 100 kHz modulation. 20 µL of every GB1 sample, each with a concentration around 30 µM, were filled into capillary tubes and the CW spectra were recorded with a field sweep of 150 G, a center field of 3505 G, and a modulation amplitude of 0.7 G. All spectra were recorded with a receiver gain of 70 dB and 1 mW power (23 dB attenuation). The double integral of each spectrum was compared with the one of a standard TEMPO sample at 100 µM concentration in water, retrieving the labelling efficiency of each nitroxide spin label for both GB1 constructs (Fig. S2, Table S2). High labelling efficiency, around 100%, was consistently observed for both GB1 constructs and for all the four distinct nitroxide labels after the overnight incubation (with the exception of MTSL for the 6C GB1 construct) as corroborated both by CW EPR and ESI-MS data. Some of the labels, especially IDSL, manifested a labelling efficiency above 100%. However, due to the low concentration of the samples, these values can be considered within the error range.

2.2 PDS Sample Preparation

Spin-labelled samples were freeze-dried and resuspended in D_2O (Merck). CuNTA stock solution (10 mM in phosphate buffer) was added to each sample to yield a final concentration of 10 µM CuNTA for the I6C/K28H/Q32H construct and 50 µM CuNTA for the I6H/N8H/K28C GB1 construct to ensure a final loading of the dHis motif of above 95%. The difference in the desired final CuNTA concentration is related to the lower affinity of the copper ligand towards the dHis motif in the β-sheet with respect to the dHis in the α-helix, as discussed previously [19]. Each sample was prepared to a final protein concentration of 10 µM and a total final volume of 65 µL, with 50% (v/v) of deuterated cryo-protectant (ethylene glycol d-6, Deutero GmbH) to ensure the formation of a glassy-frozen solution. All samples were transferred to 3 mm quartz tubes (Technical Glass Products) and immediately frozen by immersion in liquid nitrogen.

2.3 RIDME Measurements

All PDS experiments were performed using a Bruker ELEXSYS 580 pulse EPR spectrometer. Temperatures were maintained using a cryogen-free variable temperature cryostat (Cryogenic Ltd) operating in the 1.8–300 K temperature range. All

samples were measured with the 5-pulse constant-time (ct) RIDME [53] and the recently introduced variable-time (vt) RIDME [57] at 30 K, using a high-power 150 W travelling-wave tube (TWT; Applied Systems Engineering) at Q-band (34 GHz) in an overcoupled 3 mm cylindrical resonator (Bruker ER 5106QT-2w in TE102 mode). For every sample, the pulses were applied on the maximum of the nitroxide echo detected field sweep, ctRIDME measurements were performed using the pulse sequence $\pi/2$–τ_1–π–(τ_1+t)–$\pi/2$–T_{mix}–$\pi/2$–(τ_2-t)–π–τ_2–echo with detection pulse lengths $\pi/2$ and π, respectively of 8 and 16 ns. Each trace was acquired using an SRT of 10 ms, a τ_1 of 400 ns, and 2 shots-per-loop and 32-step phase cycling.

vtRIDME measurements were performed using the pulse sequence $\pi/2$–τ_1–π–(τ_1+t)–$\pi/2$–T_{mix}–$\pi/2$–τ_0–π–(τ_2+t)–echo with detection pulse lengths $\pi/2$ and π, respectively, of 8 and 16 ns. Each trace was acquired using an SRT of 10 ms, a τ_1 of 400 ns, and 2 shots-per-loop and 32-step phase cycling.

Measurements were recorded with a short (reference) and a long mixing time of 5 and 200 µs, respectively, to allow deconvolution (dividing the constant and variable-time RIDME traces with the longer mixing times by the corresponding reference traces) of the traces [57].

2.4 PDS Data Processing

Data were processed in DeerAnalysis2021 using Tikhonov regularisation and validation as previously described [58, 59]. Briefly, RIDME data were first background-corrected using a homogeneous six-dimensional background function before Tikhonov regularisation followed by statistical analysis varying background start time from 5 to 30% of the total trace length in 8 trials and varying the background dimension from 3 to 6 in 7 trials. Resulting background start time and dimension for the best fit were then used as starting points for a second round of Tikhonov regularization followed by a second round of statistical analysis, this time also including the addition of 50% random noise in 16 trials (896 total trials). Validation trials from the second validation round were pruned with a prune level of 1.15, where trials exceeding the root-mean-square deviation of the best fit by at least 15% are discarded.

2.5 Structure Prediction and Modelling

Four different GB1 structures were compared in this study: a crystallographic structure (PDB: 4wh4) and three predicted structural models obtained from AlphaFold2, OmegaFold, and ESMFold, respectively. The predicted structures were obtained through the use of Google Colabfold [60–63]. The I6C/K28H/Q32H and I6H/N8H/K28C constructs were in silico labelled using the 2021 version of MMM and a version of MtsslWizard with the newly introduced CuNTA labelling implementation [48, 49] based on octahedral coordination of the Cu^{II}. Briefly, a model of the bipedal label is superimposed onto the first labelling site and the algorithm generates 8000 trial conformations. For each of these conformations, a penalty in the form of the root-mean-square deviation (rmsd) between the main chain atoms of the second labelling site and the corresponding atoms of the label is calculated. This penalty

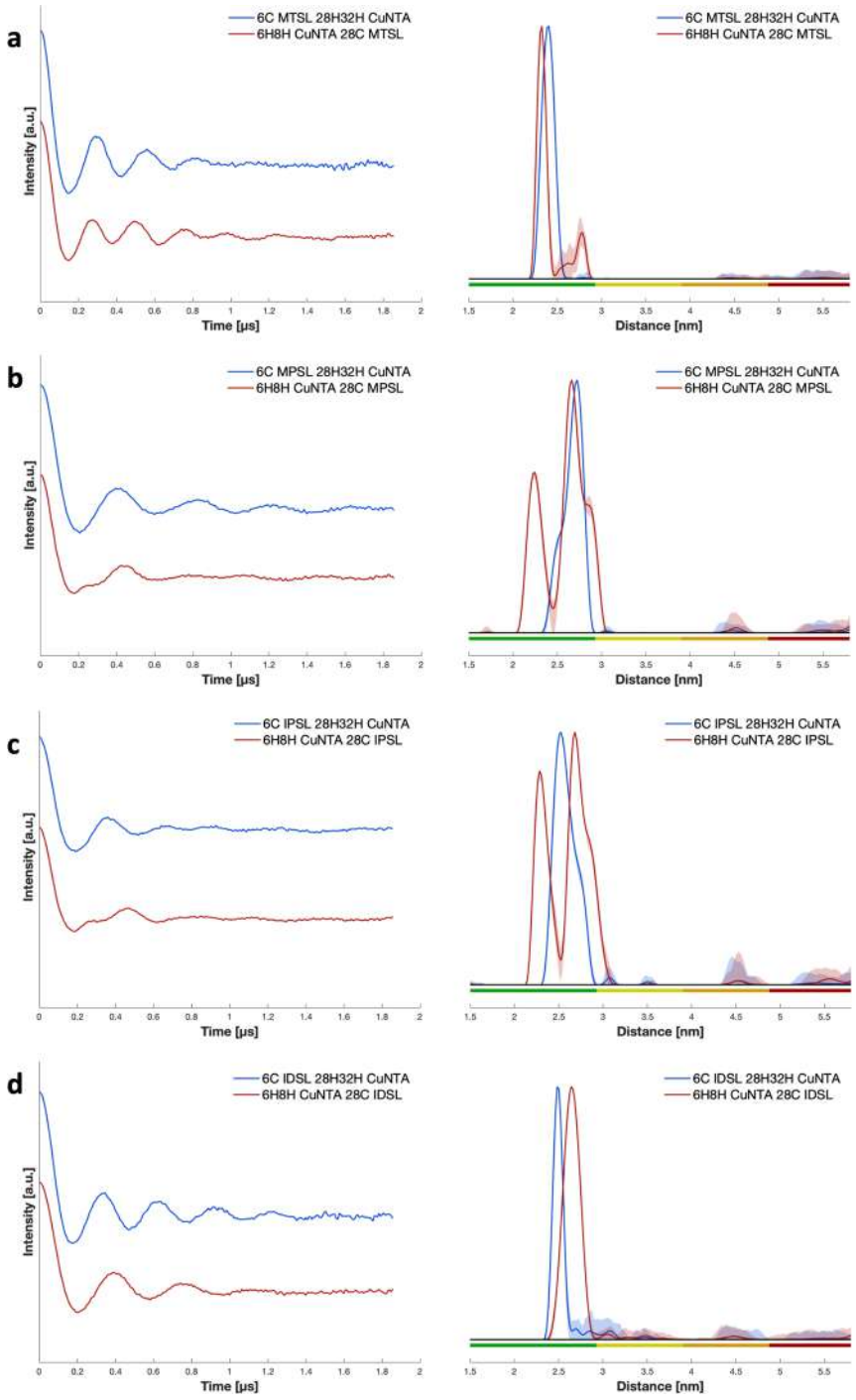
Fig. 2 Constant time RIDME traces after background correction with DeerAnalysis2021, for both GB1 constructs I6C/K28H/Q32H (nitroxide on the β-sheet, in blue) and the I6H/N8H/K28C (nitroxide on the α-helix, in red) for the four nitroxide labels, MTSL, MPSL, IPSL, and IDSL, and their respective distance distributions. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable; red: no quantification possible). **a** MTSL, **b** MPSL, **c** IPSL, and **d** IDSL

is used to rank the trial conformations and eliminate the 5% worst conformations. The discarded conformations are replaced by offspring of the remaining conformations and the process is repeated. Both labelling methods were carried out under two different conditions. Modelling with MMM was performed under ambient (298 K) and cryogenic (175 K) temperatures, and labelling with MtsslWizard was carried out with Tight (vdW-restraint 0 clashes, 3.4 Å cutoff) and Loose (vdW-restraint 15 clashes, 2.0 Å cutoff) settings [45].

3 Results and Discussion

PDS measurements were performed on the orthogonally spin-labelled GB1 constructs using the 5-pulse constant-time RIDME and the 5-pulse variable-time RIDME experiment, resulting in four different data sets (reference corrected and uncorrected for each, constant- and variable-time RIDME, Figs. S3 and S4) for each of the eight analyzed samples (two GB1 constructs with four nitroxide labels each). As expected for short distances and relatively narrow distance distributions, the experimental data did not display significant discrepancies between these four datasets, showing high consistency and robustness in the distance distributions for both ctRIDME and vtRIDME, regardless of the application or not of the deconvolution step (Fig. S5). In addition, for each of the four traces of every sample, the sensitivity per echo and per unit of time was obtained as previously described [58]. The extracted values confirmed that the vtRIDME gave generally higher sensitivity values than the ctRIDME (for more detailed information, see Tables S3 and S4).

The distributions of the experimental data of the GB1 with the nitroxide on the β-sheet site (Fig. 2) were characterized by monomodal and relatively narrow distributions, with hints of a shoulder for the MPSL and IPSL labels. On the other hand, most nitroxide labels on the α-helix displayed a bimodal trend with two distinct and large amplitude peaks arising from the presence of two different sub-ensembles of conformers. These may be rationalised for nitroxides with long enough linkers and discrete flexibility, allowing the label to be pointing away from the protein backbone at the maximum distance possible, but also allowing it to be bent towards the protein surface, resulting in shorter distances. Therefore, having the shortest and least flexible linker, IDSL can be hypothesized to have a low propensity of bending itself towards the protein surface, and was the only spin probe retaining a unimodal and relatively narrow distribution, independent of the site of GB1 to which it was attached. The observed bimodality is very unlikely caused by conformations of GB1 itself as both, the copper(II)–copper(II) distance in GB1 I6H/N8H/K28H/Q32H and


a
6C MTSL 28H32H CuNTA
6H8H CuNTA 28C MTSL
Intensity [a.u.]
Time [µs]
Distance [nm]
b
6C MPSL 28H32H CuNTA
6H8H CuNTA 28C MPSL
c
6C IPSL 28H32H CuNTA
6H8H CuNTA 28C IPSL
d
6C IDSL 28H32H CuNTA
6H8H CuNTA 28C IDSL

the nitroxide-nitroxide distance in bis-IDSL labelled GB1 6C 28C are narrow and monomodal, thus strongly indicating a single GB1 conformation [20, 26].

The presence of a spin probe with low conformational flexibility, such as dHis-CuNTA, was crucial to unravel the individual contributions of the nitroxides on the α-helix and the β-sheet. Drastically reducing the contribution to the width of one labelling site to the distance distribution allowed us to untangle the ambiguity observed in the double cysteine GB1 construct I6C/K28C [26] (Fig. S6). Surprisingly, the helix site, as a well-studied structural element which is often the preferred choice for attaching paramagnetic moieties, seemed to be the one entailing the bimodality and broadness trend of the distance distribution peaks for the system under investigation. Nonetheless, the intrinsic rigidity exhibited by the CuNTA label does not translate in discernible orientation selection effects that would impact the distance distribution shape, since in the RIDME pulse sequence, the Cu(II) is not excited by any pulses and relaxes isotropically to good approximation.

Here, we have introduced a new implementation for MtsslWizard, which can now predict distance distributions between nitroxides and dHis copper(II) labels, and we compared its performance with the ones achieved from MMM [44], to test whether, in this case, the experimental bimodality could be captured reproducibly by either or both of the in silico tools.

The in silico labelling was performed on four different model structures: the crystallographic structure (PDB: 4wh4) and three structures generated from the prediction suites AlphaFold2, OmegaFold, and ESMFold. Interestingly, a recent publication on GB1 claimed that in their research, AlphaFold2 was the best structural model in the prediction of the experimental behaviour of a dHis GB1 system [64]. While this seems counterintuitive, it may well be that the AI-based prediction performs best as it has been trained using a large number of datasets, thereby averaging many structures. Therefore, here, we discuss in depth the results obtained based on the AlphaFold2 structure, while data for the other models can be found in the SI (Tables S8–S11, Fig. S7–S10). Since no substantial variations in terms of shape, mean, width, and full width at half maximum (FWHM) of the distance distributions obtained from the vt or ct RIDME were observed, we decided to rely on the standard ctRIDME deconvoluted experimental data to compare with the in silico distance distributions. Deconvoluted data for vtRIDME and non-deconvoluted for ctRIDME and vtRIDME are shown in the SI (Fig. S4 and Tables S5–S7).

The comparison between the in silico and the experimental distance distributions (Fig. 3) revealed an intriguing difference between MMM and MtsslWizard. While the former consistently underestimated the distances between the two labels, with respect to the experimental data, the latter tended to overestimate the same measurements. In general, both approaches seemed more accurate in the prediction of the narrow and unimodal distributions of the nitroxide labelling on the β-sheet, with the only exception being MMM at cryogenic temperature settings, that simulated hints of bimodality for MPSL. On the other hand, the predicted distributions for nitroxide labelling on the α-helix were characterized by an increased broadness that partially covered the two distinct peaks of the experimental data, although neither MMM nor MtsslWizard seemed to fully and reliably predict the bimodality of the experimental data. Curiously, for MTSL, the shorter distance between the labels, corresponding

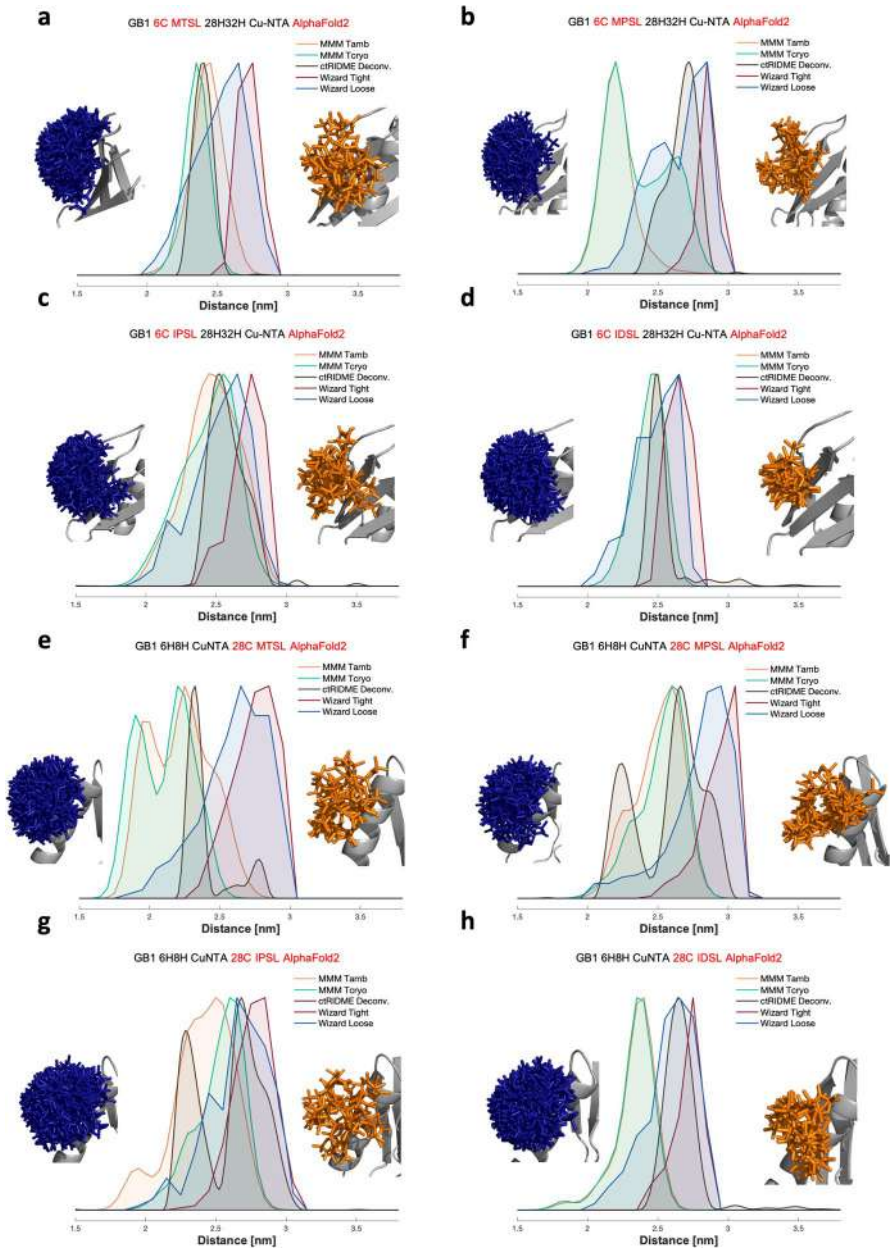


Fig. 3 Modelled distance distributions for both GB1 constructs (I6C/K28H/Q32H and I6H/N8H/K28C), both labelled with MTSL, MPSL, IPSL, and IDSL, based on the AlphaFold2 structure, superimposed with their respective experimental distance distributions (derived from ctRIDME deconvoluted data, in black). The in silico approaches compared are MMM at ambient (orange) and cryogenic (green) temperature and MtsslWizard with Tight (red) and Loose (blue) settings**.** Next to the distance distributions label rotamers modelled using MtsslWizard Loose (blue) and MMM at ambient temperature (orange) are given. **a** 6C MTSL, **b** 6C MPSL, **c** 6C IPSL, **d** 6C IDSL, **e** 28C MTSL, **f** 28C MPSL, **g** 28C IPSL, and **h** 28C IDSL

to the nitroxides interacting with the protein surface, seemed to be better predicted by MMM, while the longer distances seemed to be better predicted by MtsslWizard. This seemed sensible, as MtsslWizard has no attractive energy contribution for conformers close to the protein surface. Rotamers for each nitroxide and for the CuNTA, computed by the different labelling approaches and different conditions, are reported in Figs. S11 and S12.

To better understand and visualize the performance of the in silico approaches and of the different structure prediction tools, informative correlation plots (Fig. S13) were obtained by extracting the mean and width of every distance distribution from the experimental and the in silico labelling data (Table S5). We investigated the dependence of the labelling approach (MtsslWizard Tight or Loose and MMM ambient or cryogenic temperature) on the four different nitroxide labels. Additionally, to get a numerical quantification of the curve discrepancies, the difference between the experimental and in silico mean values was also extracted (Δ mean) (Tables S8–S11).

Altogether, neither MMM nor MtsslWizard clearly outperformed the other. For the I6C/K28H/Q32H construct, independent of the prediction method (AlphaFold2, OmegaFold, ESMFold, and X-ray structure), MTSL was predicted best by MMM at ambient temperature, while all other nitroxides were best predicted by MtsslWizard Loose. On the other hand, the interpretation of the data for the I6H/N8H/K28C construct exhibiting bimodality in the experimental distance distributions was not as straightforward. Surprisingly, MTSL, that is widely employed in PDS experiments and whose behaviour might be expected to be well known, was the label on which MMM and MtsslWizard disagreed the most and neither could satisfyingly predict the experimental mean value. This may be a consequence of failing to predict the bimodal distance distribution obtained experimentally. However, the same problem was not observed for MPSL and IPSL; while both of them yielded bimodal distributions experimentally, which again were not satisfyingly predicted by either of the two labelling approaches, the predicted mean and width values were close to the experimental ones for both MMM and MtsslWizard.

Additional correlation plots (Fig. S14) were created to evaluate how the prediction tool affected the in silico labelling performances. In general, predicting the distance distributions between the CuNTA and the nitroxide on the β-sheet site seemed to be mostly unaffected by the choice of the structure prediction method (AlphaFold2, OmegaFold, and ESMFold), as expected for a small, well-known globular protein, like GB1. In contrast, the in silico spin labelling approach clearly had a greater impact on predicted distance distributions. One exception was the prediction for IDSL estimated by MMM both at ambient and cryogenic temperature for OmegaFold, which performed substantially worse than the other prediction tools and was the furthest away from the experimental values. However, the situation was found to be more complex for the I6H/N8H/K28C construct, where the choice of the model prediction influenced the outcome of the distance distribution more strongly, although always with a lower impact than the influence derived from the labelling choice. In general, MtsslWizard seems to be less affected by this trend than MMM.

To further assess the discrepancies between the experimental and the in silico data, we extracted the rmsd values between different distributions, to quantitatively

investigate the similarity of the experimental and simulated distance distributions in terms of their overall shape (Tables S12–S15). These values mostly confirmed what was already discussed for the two modelling approaches and were highly consistent with the Δ mean values, where some values seemed to be in a better agreement with MMM, and others with MtsslWizard. Even if the bimodality was not captured by either of the two models, the increased broadness of the distributions was enough to reveal relatively low rmsd values even for the α-helix.

To comprehensively evaluate the in silico labelling methods, global rmsd values (Table S16) were extracted comparing all the mean values of the distance distributions of both GB1 constructs with the four nitroxide labels for a single labelling approach with the corresponding experimental values. In general, MtsslWizard with the Loose settings appeared to better represent the experimental data, while the other three approaches showed relatively small differences between each other. The influence of the structure prediction method was small. For MtsslWizard, the closest agreement was found using "Tight" settings based on the experimental structure, and this rmsd was significantly lower than for all other modelling approaches. Interestingly, the lowest rmsd values for MMM were obtained using the ESMFold model, whereas for both settings in MtsslWizard, the experimental structure yielded the best agreement. Very similar results could be observed for the distribution widths.

Although helices are often selected for SDSL, considering that they are well-studied secondary structural elements, in the case of GB1, it seemed to be a challenging site with respect to the sheet, giving rise to more ambiguous results in the distance distributions both from the experimental and in silico point of view. For the former, all labels, except for IDSL, showed some degree of bimodality. For the latter, we could get clues from the lower agreement between the different structure prediction methods and from the simulated distance distributions that were not able to predict the bimodality but only a higher broadness.

Overall, MMM and MtsslWizard rather complement than outperform each other. The introduced implementation for bipedal chelator ligands in the MtsslWizard now also allows predicting distance distributions between the copper and nitroxide spin labels similar to MMM. Therefore, we suggest that both approaches should be employed simultaneously to predict and compare distance distributions at a given labelling site. When the two labelling methods show significant disagreement, this could be interpreted as a warning and caution should be taken in the selection of that specific residue for experimental SDSL approaches, because the discrepancy could indicate propensity for conformational ambiguity in the PDS data.

4 Conclusions

In summary, we presented a systematic investigation of two GB1 constructs with copper(II)-nitroxide orthogonal spin labelling, alternating both labels between the α-helix and the β-sheet. We confirmed that exploiting one rigid bipedal chelator agent such as CuNTA and a nitroxide radical provides a higher precision in the distance distributions with respect to the ones obtained with two nitroxide labels. This allowed an in-depth analysis of the behaviours of different nitroxide labels for the

different sites, demonstrating how nitroxide labelling of the α-helix could introduce some ambiguity.

The flexibility of the label plays a key role regarding the width and shape of the resulting distance distribution. The chosen nitroxide should be selected carefully, considering also other properties. It is important to keep in mind that the width, mean, and shape of the distance distributions can originate from the properties of the spin label or conformational flexibility of the labelled system.

MPSL and IPSL showed the greatest differences between the α-helix and β-sheet sites; similar differences were less evident but still present for the MTSL label. In contrast, IDSL did not seem affected by the secondary structure of the labelling site, indicating how the length of the linker can affect width and shape of the distribution. It provided the narrowest and most unimodal distance distributions of all labels investigated. Therefore, we suggest that IDSL is a potentially undervalued and underused label that could possibly be exploited for gaining information on small conformational changes.

We analyzed the performance of the newly introduced bipedal labelling with copper(II) chelators for MtsslWizard on prediction of in silico distance distributions between orthogonal Cu^{II} and nitroxide labels. Overall, the results are comparable to the ones obtained from MMM, and with the help of the distribution mean, width, and rmsd values, we demonstrated how the in silico labelling approaches displayed generally good but not perfect agreement in prediction of the experimental data. However, disagreement between the two methods could be an indication of the presence of an ambiguous labelling site with, for example, a bimodal conformer distribution.

Frequently, when selecting cysteine mutants for SDSL, α-helices are chosen as well-defined structural elements. Interestingly, here, we found that this site led to more ambiguity in the interpretation of the distance distributions than the β-sheet site. Further research is required to establish whether these observations for the different secondary structures can be generalized.

The research data underpinning this publication will be accessible at https://doi.org/10.17630/71f8e2e5-9f57-4160-8c32-de1b37d4c073 [65].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00723-023-01611-1.

Acknowledgements To meet institutional and research funder open access requirements, any accepted manuscript arising shall be open access under a Creative Commons Attribution (CC BY) reuse licence with zero embargo. We acknowledge support by the Wellcome Trust (204821/Z/16/Z), the Leverhulme Trust (RPG-2018-397), and the EPSRC (EP/X016455/1). BEB acknowledges equipment funding by BBSRC (BB/R013780/1 and BB/T017740/1).

Author Contributions VV: investigation, data curation, formal analysis, writing—original draft, and writing—review and editing. KA: conceptualization, investigation, data curation, formal analysis, and writing—review and editing. GH: methodology and writing—review and editing. BEB: conceptualization, investigation, funding acquisition, methodology, supervision, and writing—review and editing.

Funding The authors acknowledge support by the Wellcome Trust (204821/Z/16/Z), the Leverhulme Trust (RPG-2018-397), and the EPSRC (EP/X016455/1). BEB acknowledges equipment funding by BBSRC (BB/R013780/1 and BB/T017740/1).

Availability of Data and Materials The research data underpinning this publication will be accessible at https://doi.org/10.17630/71f8e2e5-9f57-4160-8c32-de1b37d4c073 [65].

Declarations

Conflict of Interest All authors declare that they have no conflict of interest.

Ethical Approval Not applicable.

References

1. B. Verhalen, R. Dastvan, S. Thangapandian, Y. Peskova, H.A. Koteiche, R.K. Nakamoto, E. Tajkhorshid, H.S. McHaourab, Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. Nature **543**, 738–741 (2017). https://doi.org/10.1038/nature21414
2. D.T. Edwards, T. Huber, S. Hussain, K.M. Stone, M. Kinnebrew, I. Kaminker, E. Matalon, M.S. Sherwin, D. Goldfarb, S. Han, Determining the oligomeric structure of proteorhodopsin by Gd^{3+}-based pulsed dipolar spectroscopy of multiple distances. Structure **22**, 1677–1686 (2014). https://doi.org/10.1016/j.str.2014.09.008
3. M. Kasanmascheff, W. Lee, T.U. Nick, J. Stubbe, M. Bennati, Radical transfer in *E. coli* ribonucleotide reductase: a NH_2Y_{731}/R_{411}A-α mutant unmasks a new conformation of the pathway residue 731. Chem. Sci. **7**, 2170–2178 (2016). https://doi.org/10.1039/c5sc03460d
4. O. Schiemann, N. Piton, Y. Mu, G. Stock, J.W. Engels, T.F. Prisner, A PELDOR-based nanometer distance ruler for oligonucleotides. J. Am. Chem. Soc. **126**, 5722–5729 (2004). https://doi.org/10.1021/ja0393877
5. E.S. Babaylova, A.A. Malygin, A.A. Lomzov, D.V. Pyshnyi, M. Yulikov, G. Jeschke, O.A. Krumkacheva, M.V. Fedin, G.G. Karpova, E.G. Bagryanskaya, Complementary-addressed site-directed spin labeling of long natural RNAs. Nucl. Acids Res. **44**, 7935–7943 (2016). https://doi.org/10.1093/nar/gkw516
6. Q. Cai, A.K. Kusnetzow, W.L. Hubbell, I.S. Haworth, G.P. Gacho, N. Van Eps, K. Hideg, E.J. Chambers, P.Z. Qin, Site-directed spin labeling measurements of nanometer distances in nucleic acids using a sequence-independent nitroxide probe. Nucl. Acids Res. **34**, 4722–4730 (2006). https://doi.org/10.1093/nar/gkl546
7. C. Altenbach, S.L. Flitsch, H.G. Khorana, W.L. Hubbell, Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. Biochemistry **28**, 7806–7812 (1989). https://doi.org/10.1021/bi00445a042
8. C. Altenbach, T. Marti, H.G. Khorana, W.L. Hubbell, Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. Science **248**, 1088–1092 (1990). https://doi.org/10.1126/science.2160734
9. C. Altenbach, W. Froncisz, J.S. Hyde, W.L. Hubbell, Conformation of spin-labeled melittin at membrane surfaces investigated by pulse saturation recovery and continuous wave power saturation electron paramagnetic resonance. Biophys. J. **56**, 1183–1191 (1989). https://doi.org/10.1016/S0006-3495(89)82765-1

10. T. Schmidt, M.A. Walti, J.L. Baber, E.J. Hustedt, G.M. Clore, Long distance measurements up to 160 Å in the GroEL tetradecamer using Q-band DEER EPR spectroscopy. Angew. Chem. Int. Ed. Engl. **55**, 15905–15909 (2016). https://doi.org/10.1002/anie.201609617
11. R. Ward, A. Bowman, E. Sozudogru, H. El-Mkami, T. Owen-Hughes, D.G. Norman, EPR distance measurements in deuterated proteins. J. Magn. Reson. **207**, 164–167 (2010). https://doi.org/10.1016/j.jmr.2010.08.002
12. T.M. Thaker, S. Mishra, W. Zhou, M. Mohan, Q. Tang, J.D. Faraldo-Gomez, H.S. McHaourab, T.M. Tomasiak, Asymmetric drug binding in an ATP-loaded inward-facing state of an ABC transporter. Nat. Chem. Biol. **18**, 226–235 (2022). https://doi.org/10.1038/s41589-021-00936-x
13. P.S. Kerry, H.L. Turkington, K. Ackermann, S.A. Jameison, B.E. Bode, Analysis of influenza A virus NS1 dimer interfaces in solution by pulse EPR distance measurements. J. Phys. Chem. B **118**, 10882–10888 (2014). https://doi.org/10.1021/jp508386r
14. B. Joseph, A. Sikora, D.S. Cafiso, Ligand induced conformational changes of a membrane transporter in *E. coli* cells observed with DEER/PELDOR. J. Am. Chem. Soc. **138**, 1844–1847 (2016). https://doi.org/10.1021/jacs.5b13382
15. C. Kapsalis, B. Wang, H. El Mkami, S.J. Pitt, J.R. Schnell, T.K. Smith, J.D. Lippiat, B.E. Bode, C. Pliotas, Allosteric activation of an ion channel triggered by modification of mechanosensitive nanopockets. Nat. Commun. **10**, 4619 (2019). https://doi.org/10.1038/s41467-019-12591-x
16. H. Sameach, S. Ghosh, L. Gevorkyan-Airapetov, S. Saxena, S. Ruthstein, EPR spectroscopy detects various active state conformations of the transcriptional regulator CueR. Angew. Chem. Int. Ed. Engl. **58**, 3053–3056 (2019). https://doi.org/10.1002/anie.201810656
17. L. Galazzo, G. Meier, M.H. Timachi, C.A.J. Hutter, M.A. Seeger, E. Bordignon, Spin-labeled nanobodies as protein conformational reporters for electron paramagnetic resonance in cellular membranes. Proc. Natl. Acad. Sci. USA **117**, 2441–2448 (2020). https://doi.org/10.1073/pnas.1913737117
18. T.O. Peulen, C.S. Hengstenberg, R. Biehl, M. Dimura, C. Lorenz, A. Valeri, J. Folz, C.A. Hanke, S. Ince, T. Vopel, B. Farago, H. Gohlke, J.P. Klare, A.M. Stadler, C.A.M. Seidel, C. Herrmann, Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase. Elife (2023). https://doi.org/10.7554/eLife.79565
19. J.L. Wort, K. Ackermann, A. Giannoulis, A.J. Stewart, D.G. Norman, B.E. Bode, Sub-micromolar pulse dipolar EPR spectroscopy reveals increasing Cu^{II}-labelling of double-histidine motifs with lower temperature. Angew. Chem. Int. Ed. Engl. **58**, 11681–11685 (2019). https://doi.org/10.1002/anie.201904848
20. T.F. Cunningham, M.R. Putterman, A. Desai, W.S. Horne, S. Saxena, The double-histidine Cu^{2+}-binding motif: a highly rigid, site-specific spin probe for electron spin resonance distance measurements. Angew. Chem. Int. Ed. Engl. **54**, 6330–6334 (2015). https://doi.org/10.1002/anie.201501968
21. O. Schiemann, T.F. Prisner, Long-range distance determinations in biomacromolecules by EPR spectroscopy. Q. Rev. Biophys. **40**, 1–53 (2007). https://doi.org/10.1017/S003358350700460X
22. A. Giannoulis, A. Feintuch, Y. Barak, H. Mazal, S. Albeck, T. Unger, F. Yang, X.C. Su, D. Goldfarb, Two closed ATP- and ADP-dependent conformations in yeast Hsp90 chaperone detected by Mn(II) EPR spectroscopic techniques. Proc. Natl. Acad. Sci. USA **117**, 395–404 (2020). https://doi.org/10.1073/pnas.1916030116
23. H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb, G. Otting, Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. J. Am. Chem. Soc. **133**, 10418–10421 (2011). https://doi.org/10.1021/ja204415w
24. A. Shah, A. Roux, M. Starck, J.A. Mosely, M. Stevens, D.G. Norman, R.I. Hunter, H. El Mkami, G.M. Smith, D. Parker, J.E. Lovett, A Gadolinium spin label with both a narrow central transition and short tether for use in double electron electron resonance distance measurements. Inorg. Chem. **58**, 3015–3025 (2019). https://doi.org/10.1021/acs.inorgchem.8b02892
25. K.H. Richardson, J.J. Wright, M. Simenas, J. Thiemann, A.M. Esteves, G. McGuire, W.K. Myers, J.J.L. Morton, M. Hippler, M.M. Nowaczyk, G.T. Hanke, M.M. Roessler, Functional basis of electron transport within photosynthetic complex I. Nat. Commun. **12**, 5387 (2021). https://doi.org/10.1038/s41467-021-25527-1
26. K. Ackermann, A. Chapman, B.E. Bode, A comparison of cysteine-conjugated nitroxide spin labels for pulse dipolar EPR spectroscopy. Molecules (2021). https://doi.org/10.3390/molecules26247534
27. G. Jeschke, DEER distance measurements on proteins. Annu. Rev. Phys. Chem. **63**, 419–446 (2012). https://doi.org/10.1146/annurev-physchem-032511-143716

28. L.J. Berliner, J. Grunwald, H.O. Hankovszky, K. Hideg, A novel reversible thiol-specific spin label: papain active site labeling and inhibition. Anal. Biochem. **119**, 450–455 (1982). https://doi.org/10.1016/0003-2697(82)90612-1
29. T. Pali, M.E. Finbow, D. Marsh, Membrane assembly of the 16-kDa proteolipid channel from *Nephrops norvegicus* studied by relaxation enhancements in spin-label ESR. Biochemistry **38**, 14311–14319 (1999). https://doi.org/10.1021/bi991459c
30. Y. Shenberger, H.E. Gottlieb, S. Ruthstein, EPR and NMR spectroscopies provide input on the coordination of Cu(I) and Ag(I) to a disordered methionine segment. J. Biol. Inorg. Chem. **20**, 719–727 (2015). https://doi.org/10.1007/s00775-015-1259-1
31. H. Dugas, A. Rodriguez, N. Brisson, Selective spin-labeling of the ribosomal proteins of 70S ribosomes from Escherichia coli. Can. J. Biochem. **57**, 1407–1415 (1979). https://doi.org/10.1139/o79-187
32. C. Gmeiner, G. Dorn, F.H.T. Allain, G. Jeschke, M. Yulikov, Spin labelling for integrative structure modelling: a case study of the polypyrimidine-tract binding protein 1 domains in complexes with short RNAs. Phys. Chem. Chem. Phys. **19**, 28360–28380 (2017). https://doi.org/10.1039/c7cp05822e
33. A. Volkov, C. Dockter, T. Bund, H. Paulsen, G. Jeschke, Pulsed EPR determination of water accessibility to spin-labeled amino acid residues in LHCIIb. Biophys. J. **96**, 1124–1141 (2009). https://doi.org/10.1016/j.bpj.2008.09.047
34. P.B. Hewgley, D. Puett, Spin labeled calmodulin: a new probe for studying Ca^{2+} and macromolecular interactions. Ann. N. Y. Acad. Sci. **356**, 20–32 (1980). https://doi.org/10.1111/j.1749-6632.1980.tb29595.x
35. A.R. Balo, H. Feyrer, O.P. Ernst, Toward precise interpretation of DEER-based distance distributions: insights from structural characterization of V1 spin-labeled side chains. Biochemistry **55**, 5256–5263 (2016). https://doi.org/10.1021/acs.biochem.6b00608
36. L.M. Wingler, M. Elgeti, D. Hilger, N.R. Latorraca, M.T. Lerch, D.P. Staus, R.O. Dror, B.K. Kobilka, W.L. Hubbell, R.J. Lefkowitz, Angiotensin analogs with divergent bias stabilize distinct receptor conformations. Cell **176**, 468-478 e411 (2019). https://doi.org/10.1016/j.cell.2018.12.005
37. D. Toledo Warshaviak, V.V. Khramtsov, D. Cascio, C. Altenbach, W.L. Hubbell, Structure and dynamics of an imidazoline nitroxide side chain with strongly hindered internal motion in proteins. J. Magn. Reson. **232**, 53–61 (2013). https://doi.org/10.1016/j.jmr.2013.04.013
38. V.V. Khramtsov, V.I. Yelinova, L.M. Weiner, T.A. Berezina, V.V. Martin, L.B. Volodarsky, Quantitative determination of SH groups in low- and high-molecular-weight compounds by an electron spin resonance method. Anal. Biochem. **182**, 58–63 (1989). https://doi.org/10.1016/0003-2697(89)90718-5
39. S. Ghosh, M.J. Lawless, G.S. Rule, S. Saxena, The Cu^{2+}-nitrilotriacetic acid complex improves loading of alpha-helical double histidine site for precise distance measurements by pulsed ESR. J. Magn. Reson. **286**, 163–171 (2018). https://doi.org/10.1016/j.jmr.2017.12.005
40. M.J. Lawless, S. Ghosh, T.F. Cunningham, A. Shimshi, S. Saxena, On the use of the Cu^{2+}-iminodiacetic acid complex for double histidine based distance measurements by pulsed ESR. Phys. Chem. Chem. Phys. **19**, 20959–20967 (2017). https://doi.org/10.1039/c7cp02564e
41. J. Voss, L. Salwinski, H.R. Kaback, W.L. Hubbell, A method for distance determination in proteins using a designed metal ion binding site and site-directed spin labeling: evaluation with T4 lysozyme. Proc. Natl. Acad. Sci. USA **92**, 12295–12299 (1995). https://doi.org/10.1073/pnas.92.26.12295
42. J.L. Wort, S. Arya, K. Ackermann, A.J. Stewart, B.E. Bode, Pulse dipolar EPR reveals double-histidine motif Cu^{II}-NTA spin-labeling robustness against competitor ions. J. Phys. Chem. Lett. **12**, 2815–2819 (2021). https://doi.org/10.1021/acs.jpclett.1c00211
43. G. Jeschke, MMM: a toolbox for integrative structure modeling. Protein Sci. **27**, 76–85 (2018). https://doi.org/10.1002/pro.3269
44. S. Ghosh, S. Saxena, G. Jeschke, Rotamer modelling of Cu(II) spin labels based on the double-histidine motif. Appl. Magn. Reson. **49**, 1281–1298 (2018). https://doi.org/10.1007/s00723-018-1052-8
45. G. Hagelueken, R. Ward, J.H. Naismith, O. Schiemann, MtsslWizard: in silico spin-labeling and generation of distance distributions in PyMOL. Appl. Magn. Reson. **42**, 377–391 (2012). https://doi.org/10.1007/s00723-012-0314-0
46. M.H. Tessmer, S. Stoll, ChiLife: an open-source Python package for in silico spin labeling and integrative protein modeling. PLoS Comput. Biol. (2023). https://doi.org/10.1371/journal.pcbi.1010834

47. M.H. Tessmer, S. Stoll, A rotamer library approach to modeling side chain ensembles of the bifunctional spin label RX. Appl. Magn. Reson. (2023). https://doi.org/10.1007/s00723-023-01576-1
48. G. Hagelueken. mtsslWizard for bipedal labels. 2023. https://github.com/gha2012/mtsslWizard_bipedal
49. G. Hagelueken. Colab running mtsslWizard for bipedal labels. 2023. https://github.com/gha2012/mtsslWizard_colab/blob/main/mtsslWizard_colab.ipynb
50. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. Nature **596**, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
51. R. Wu, F. Ding, R. Wang, R. Shen, X. Zhang, S. Luo, C. Su, Z. Wu, Q. Xie, B. Berger, J. Ma, J. Peng, High-resolution de novo structure prediction from primary sequence. bioRxiv (2022). https://doi.org/10.1101/2022.07.21.500999
52. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**, 1123–1130 (2023). https://doi.org/10.1126/science.ade2574
53. S. Milikisyants, F. Scarpelli, M.G. Finiguerra, M. Ubbink, M. Huber, A pulsed EPR method to determine distances between paramagnetic centers with strong spectral anisotropy and radicals: the dead-time free RIDME sequence. J. Magn. Reson. **201**, 48–56 (2009). https://doi.org/10.1016/j.jmr.2009.08.008
54. L.V. Kulik, S.A. Dzuba, I.A. Grigoryev, Y.D. Tsvetkov, Electron dipole–dipole interaction in ESEEM of nitroxide biradicals. Chem. Phys. Lett. **343**, 315–324 (2001). https://doi.org/10.1016/S0009-2614(01)00721-7
55. K. Keller, M. Qi, C. Gmeiner, I. Ritsch, A. Godt, G. Jeschke, A. Savitsky, M. Yulikov, Intermolecular background decay in RIDME experiments. Phys. Chem. Chem. Phys. **21**, 8228–8245 (2019). https://doi.org/10.1039/c8cp07815g
56. A. Giannoulis, C.L. Motion, M. Oranges, M. Buhl, G.M. Smith, B.E. Bode, Orientation selection in high-field RIDME and PELDOR experiments involving low-spin Co^{II} ions. Phys. Chem. Chem. Phys. **20**, 2151–2154 (2018). https://doi.org/10.1039/c7cp07248a
57. J.L. Wort, K. Ackermann, A. Giannoulis, B.E. Bode, Enhanced sensitivity for pulse dipolar EPR spectroscopy using variable-time RIDME. J. Magn. Reson. **352**, 107460 (2023). https://doi.org/10.1016/j.jmr.2023.107460
58. K. Ackermann, J.L. Wort, B.E. Bode, Pulse dipolar EPR for determining nanomolar binding affinities. Chem. Commun. **58**, 8790–8793 (2022). https://doi.org/10.1039/d2cc02360a
59. K. Ackermann, J.L. Wort, B.E. Bode, Nanomolar pulse dipolar EPR spectroscopy in proteins: Cu^{II}–Cu^{II} and nitroxide–nitroxide cases. J. Phys. Chem. B **125**, 5358–5364 (2021). https://doi.org/10.1021/acs.jpcb.1c03666
60. Google Colabfold ESMFold. 2023. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/ESMFold.ipynb
61. Google Colabfold OmegaFold. 2023. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/omegafold.ipynb
62. Google Colabfold AlphaFold2. 2023. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
63. M. Mirdita, K. Schutze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. Nat. Methods **19**, 679 (2022). https://doi.org/10.1038/s41592-022-01488-1
64. W. Zhu, D.T. Yang, A.M. Gronenborn, Ligand-capped cobalt(II) multiplies the value of the double-histidine motif for PCS NMR studies. J. Am. Chem. Soc. **145**, 4564–4569 (2023). https://doi.org/10.1021/jacs.2c12021
65. V. Vitali, K. Ackermann, G. Hagelueken, B.E. Bode, Spectroscopically orthogonal labelling to disentangle site-specific nitroxide label distributions, in *Dataset* (2023). University of St Andrews Research Portal. https://doi.org/10.17630/71f8e2e5-9f57-4160-8c32-de1b37d4c073

Authors and Affiliations

Valentina Vitali[1,2,3] · Katrin Ackermann[1] · Gregor Hagelueken[4] · Bela E. Bode[1]

✉ Bela E. Bode
beb2@st-andrews.ac.uk

Gregor Hagelueken
hagelueken@uni-bonn.de

1 EaStCHEM School of Chemistry, Biomedical Sciences Research Complex, and Centre of Magnetic Resonance, University of St Andrews, North Haugh, St Andrews KY16 9ST, Scotland

2 Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy

3 Department of Chemistry "Ugo Schiff", University of Florence, Via Della Lastruccia 3, 50019 Sesto Fiorentino, Italy

4 Institute of Structural Biology, Biomedical Center, University of Bonn, Venusberg-Campus 1, 53127 Bonn, Germany

Supplementary Information

Spectroscopically orthogonal labelling to disentangle site-specific nitroxide label distributions

Valentina Vitali[1,2], Katrin Ackermann[1], Gregor Hagelueken*[3], and Bela E. Bode*[1]

[1]EaStCHEM School of Chemistry, Biomedical Sciences Research Complex, and Centre of Magnetic Resonance, University of St Andrews, North Haugh, St Andrews, KY16 9ST, Scotland; [2] Magnetic Resonance Center (CERM), University of Florence, via Luigi Sacconi 6, Sesto Fiorentino 50019, Italy; Department of Chemistry "Ugo Schiff", University of Florence, via della Lastruccia 3, Sesto Fiorentino 50019, Italy; [3]Institute of Structural Biology, Biomedical Center, University of Bonn, Venusberg-Campus 1, 53127 Bonn, Germany

Table of contents

I Results and Discussion

1.1 Mass Spectrometry

Acquired ESI mass spectrometry spectra show only a single peak at the expected mass of the protein (unlabelled or labelled with the four nitroxides, respectively). The only exception is for the 6C MTSL sample where a peak with smaller intensity belonging to the unreacted protein can also be detected (**Fig. S1**). For the GB1 I6H/N8H/K28C control construct, a smaller peak corresponding to the dimer (12487 Da) can be observed. However, no dimer peak was detected in any of the labelled samples, leading us to assume that its formation may have occurred during the time frame between removing the DTT and performing the mass spectrometry measurement.
The expected masses and the actual experimental masses for the different samples are reported in **Table S1**.

ESI-MS

Mass values (Da)	Control		MTSL		MPSL		IPSL		IDSL	
	6C	28C	6C	28C	6C	28C	6C	28C	6C	28C
Expected	6244.8	6230.8	6414.8	6428.8	6467.8	6481.8	6427.8	6441.8	6401.8	6415.8
Obtained	6244.0	6230.6	6414.8	6429.0	6467.9	6481.0	6427.7	6441.1	6401.8	6415.1

Table S1 Expected and obtained mass values (Da) for both GB1 I6C/K28H/Q32H and I6H/N8H/K28C GB1 constructs with and without (control) the four different nitroxide labels MTSL, MPSL, IPSL and IDSL

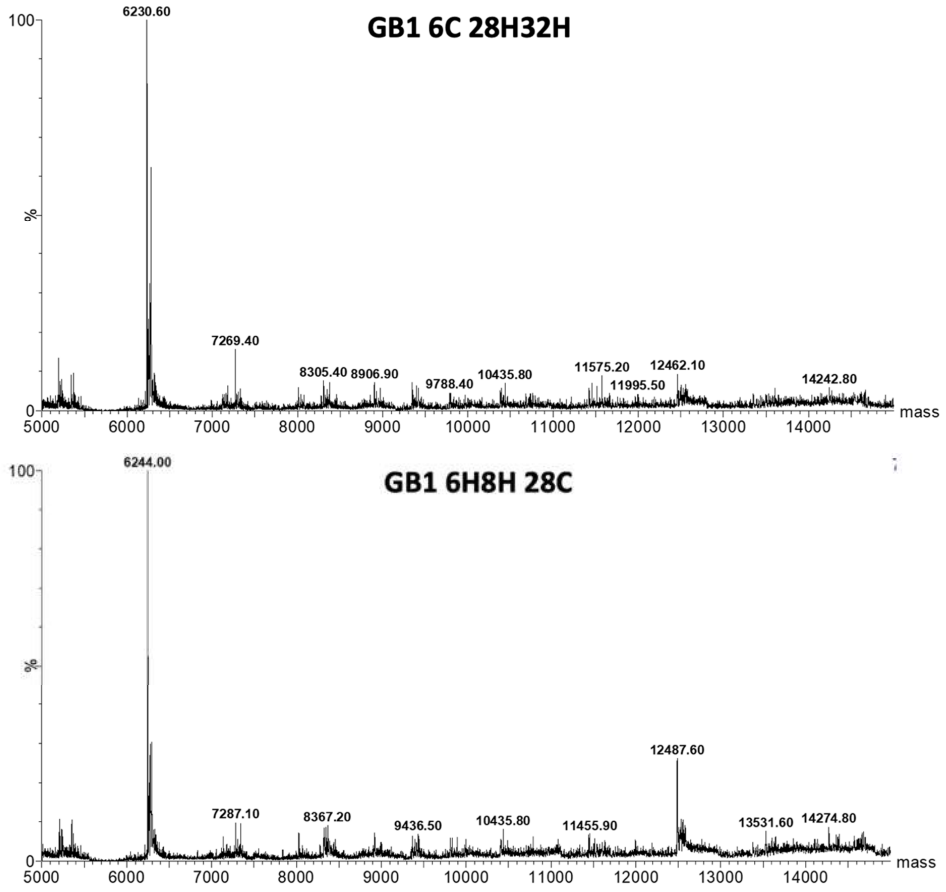


Fig. S1 ESI-MS spectra of the two GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C with and without (control) the four nitroxide labels MTSL, MPSL, IPSL and IDSL

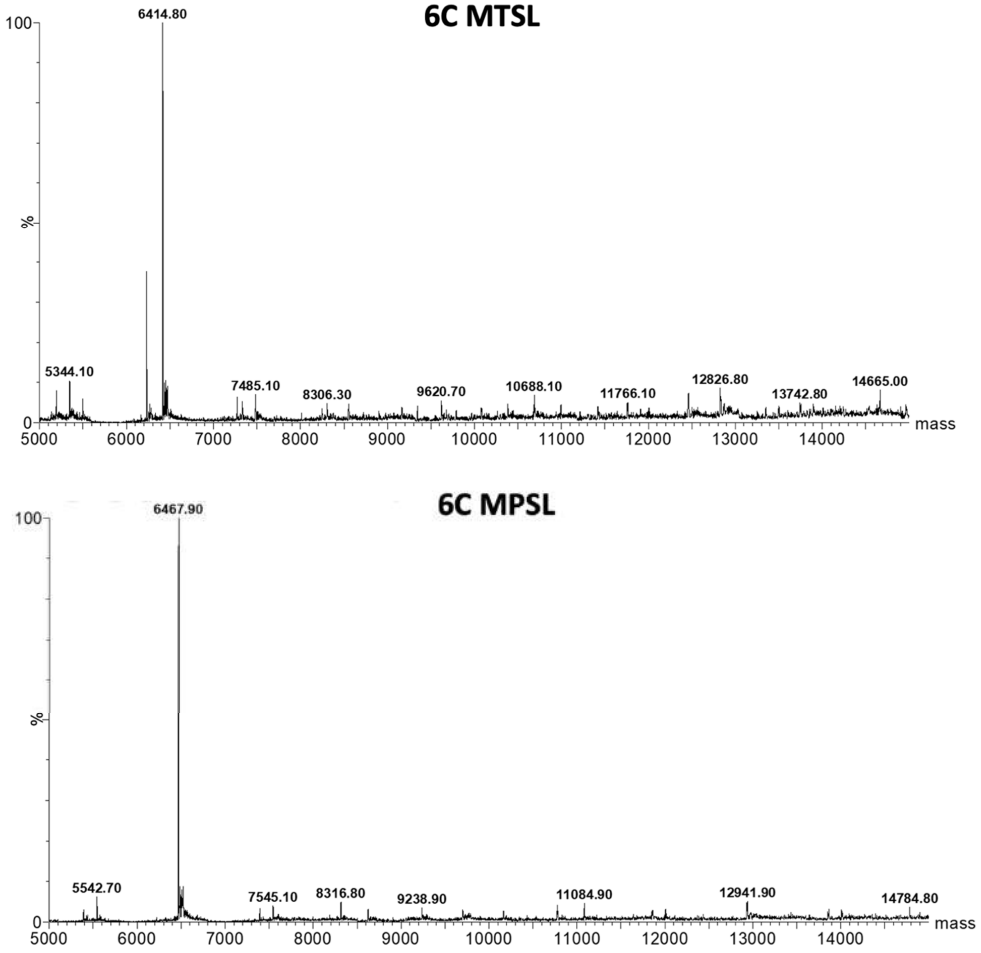


Fig. S1, continued ESI-MS spectra of the two GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C with and without (control) the four nitroxide labels MTSL, MPSL, IPSL and IDSL

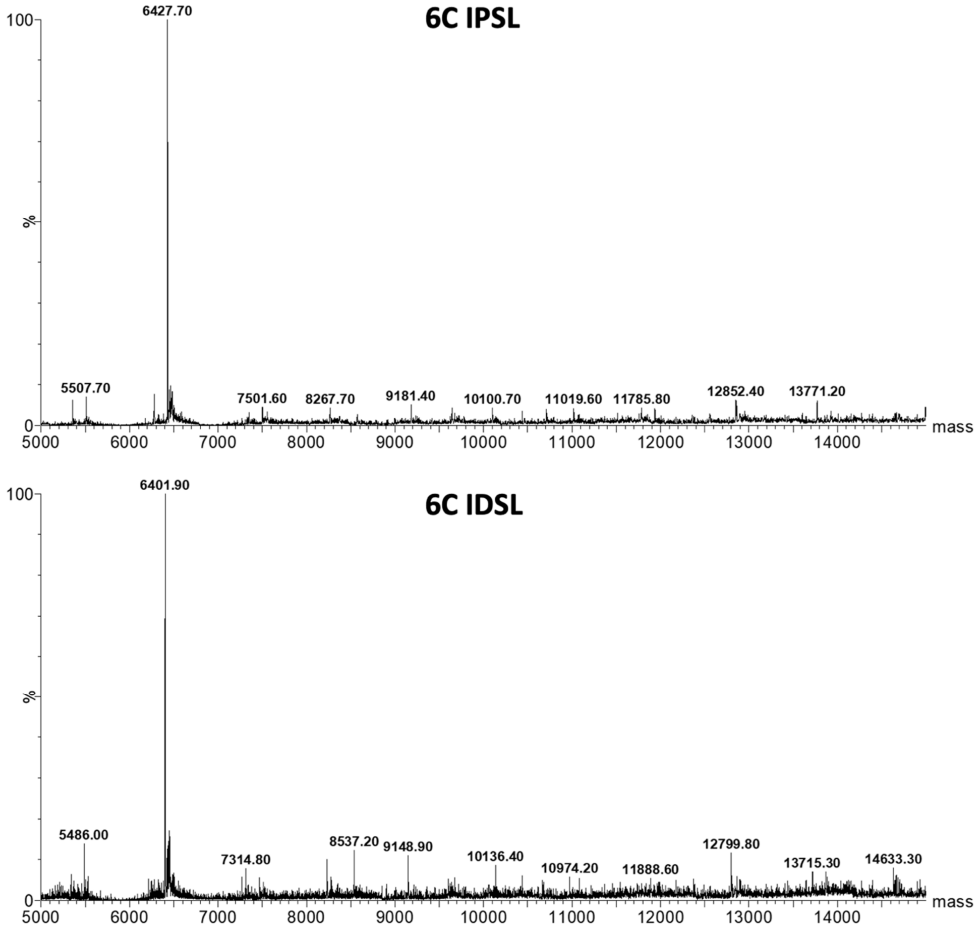


Fig. S1, continued ESI-MS spectra of the two GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C with and without (control) the four nitroxide labels MTSL, MPSL, IPSL and IDSL

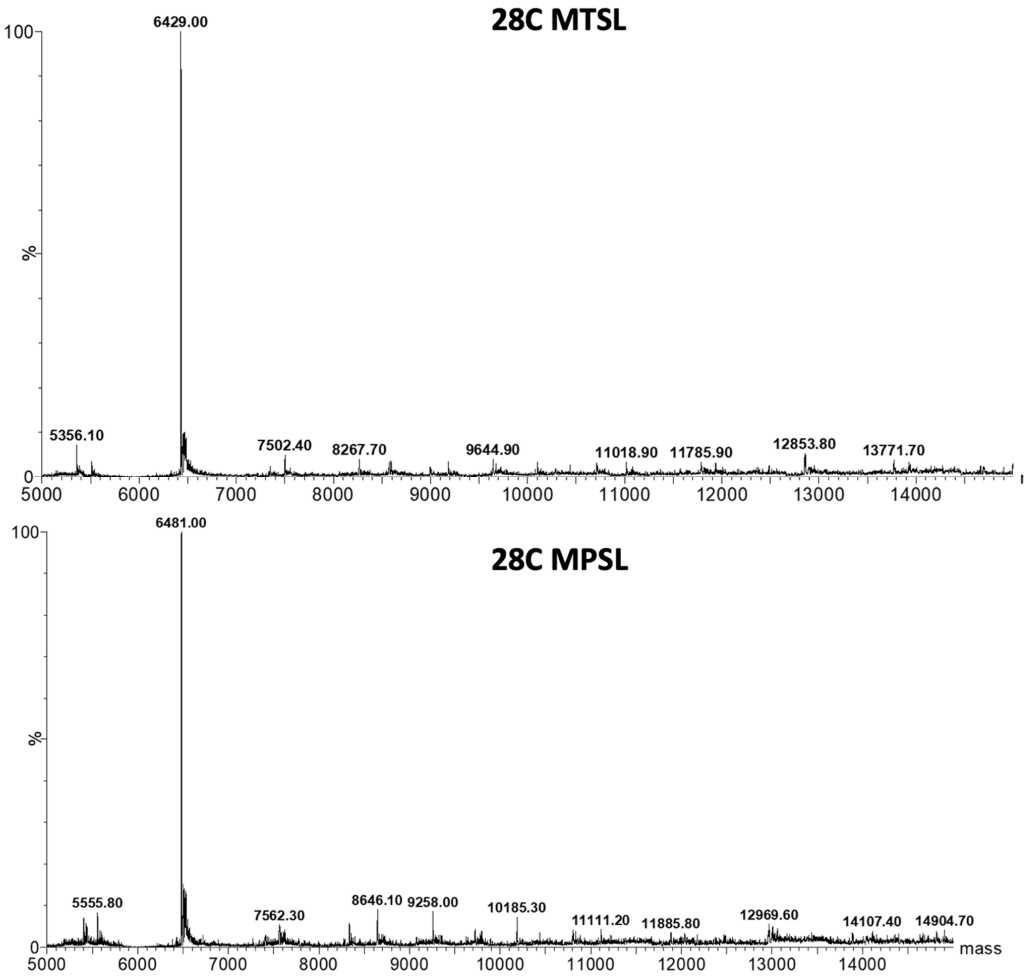


Fig. S1, continued ESI-MS spectra of the two GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C with and without (control) the four nitroxide labels MTSL, MPSL, IPSL and IDSL

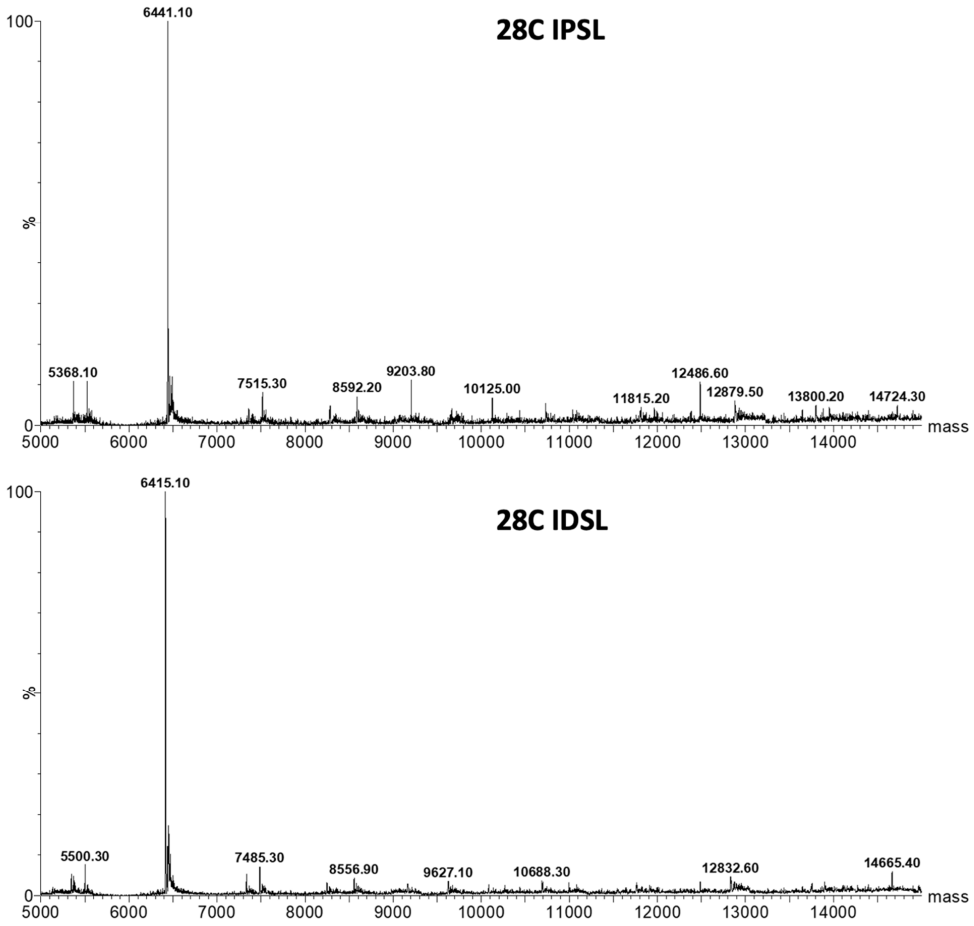


Fig. S1, continued ESI-MS spectra of the two GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C with and without (control) the four nitroxide labels MTSL, MPSL, IPSL and IDSL

1.2 Continuous Wave EPR Spectroscopy

Individual CW EPR spectra and corresponding labelling efficiencies are given in **Fig. S2** and **Table S2**. Quantitative labelling was obtained for all but the MTSL labelling of the I6C/K28H/Q32H construct, where the labelling efficiency was found to be lower, consistent with what was observed from the ESI-MS spectra.

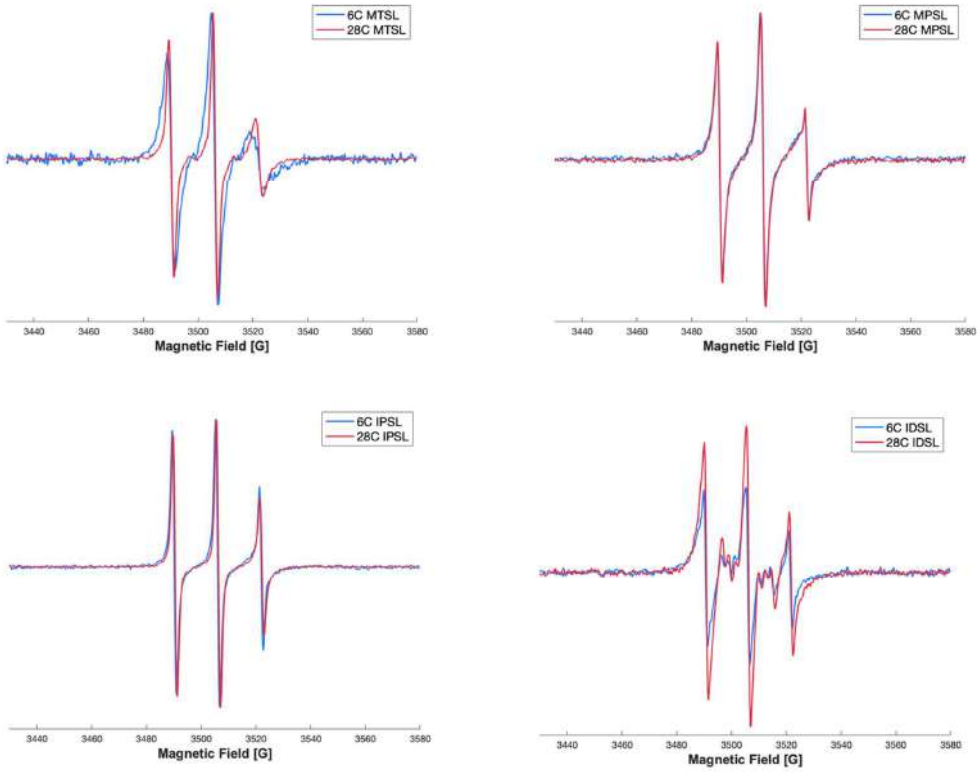


Fig. S2 CW-EPR plots of the GB1 I6C/K28H/Q32H (in blue) and I6H/N8H/K28C (in red) GB1 constructs with the four distinct nitroxide labels, MTSL, MPSL, IPSL and IDSL

Labelling Efficiency

CW-EPR	MTSL		MPSL		IPSL		IDSL	
	6C	28C	6C	28C	6C	28C	6C	28C
	87%	98%	115%	99%	105%	100%	114%	127%

Table S2 Labelling efficiencies for both GB1 I6C/K28H/Q32H and I6H/N8H/K28C GB1 constructs with the four nitroxides labels, MTSL, MPSL, IPSL and IDSL

1.3 RIDME raw spectra

RIDME raw spectra for both the variable and constant time RIDME, their reference background traces and the superimposed traces after deconvolution are reported in **Fig. S3.**

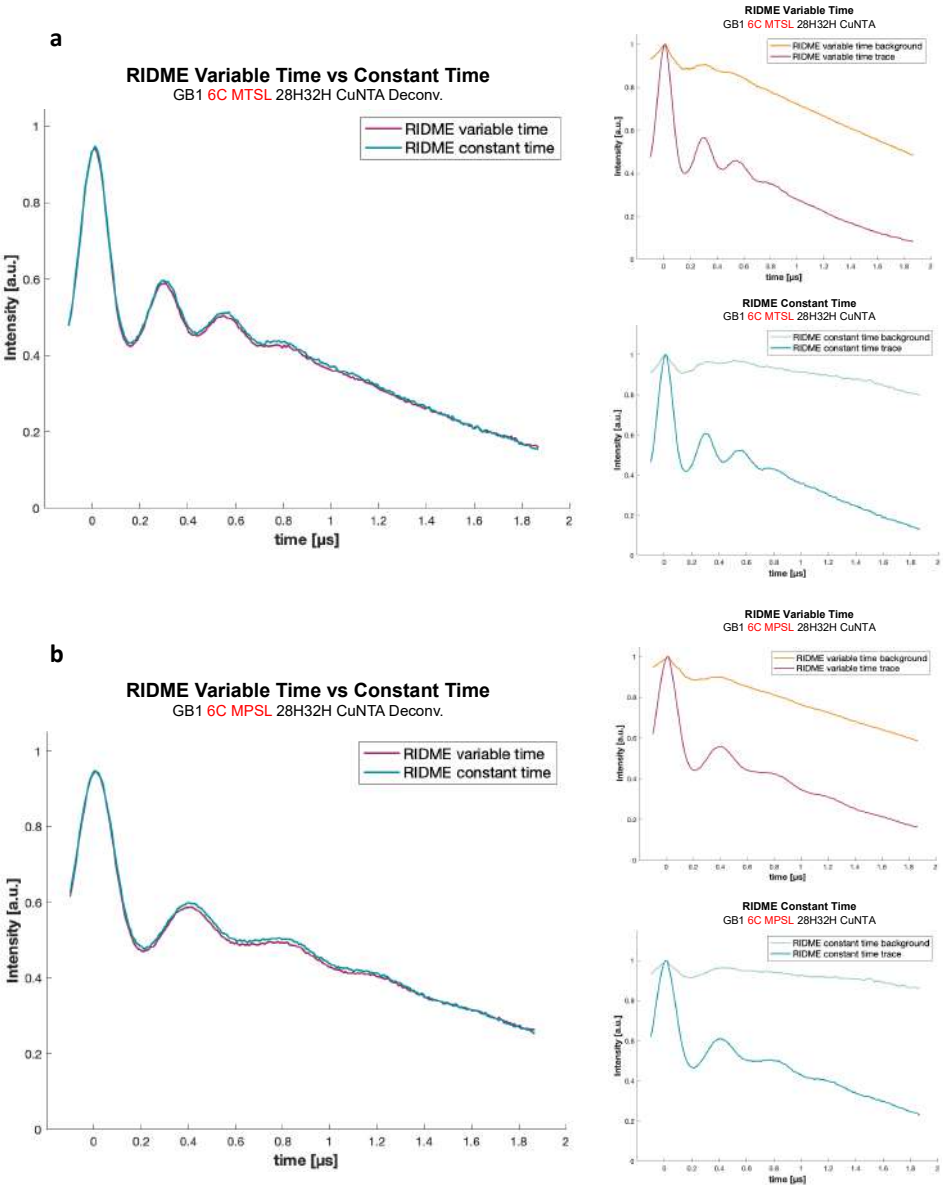


Fig. S3 Right: ctRIDME (cyan) and vtRIDME (magenta) raw data traces and their respective background traces (light green and orange). Left: superimposed constant time and variable time RIDME traces after deconvolution **a)** GB1 I6C/K28H/Q32H MTSL Cu-NTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

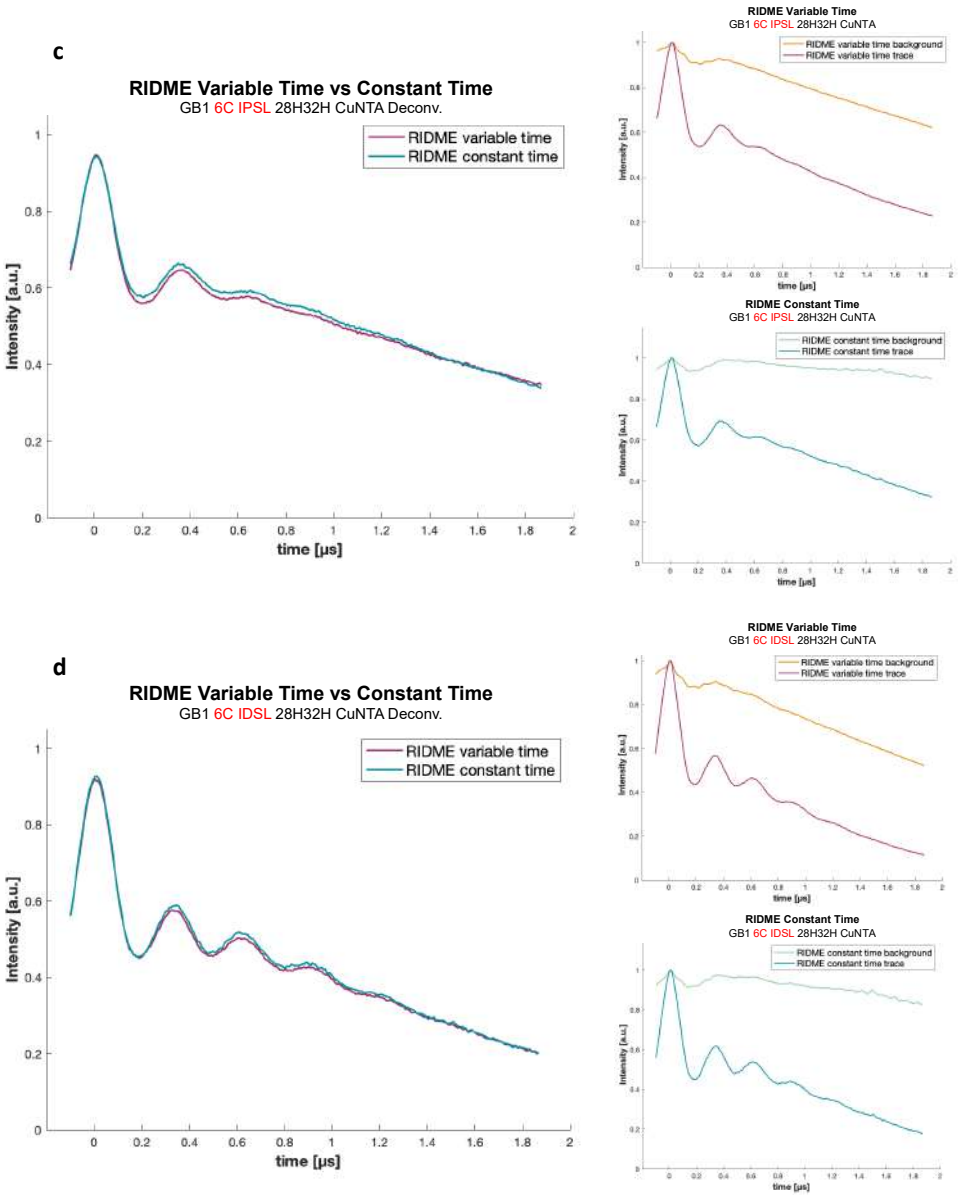


Fig. S3, continued Right: ctRIDME (cyan) and vtRIDME (magenta) raw data traces and their respective background traces (light green and orange). Left: superimposed constant time and variable time RIDME traces after deconvolution **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

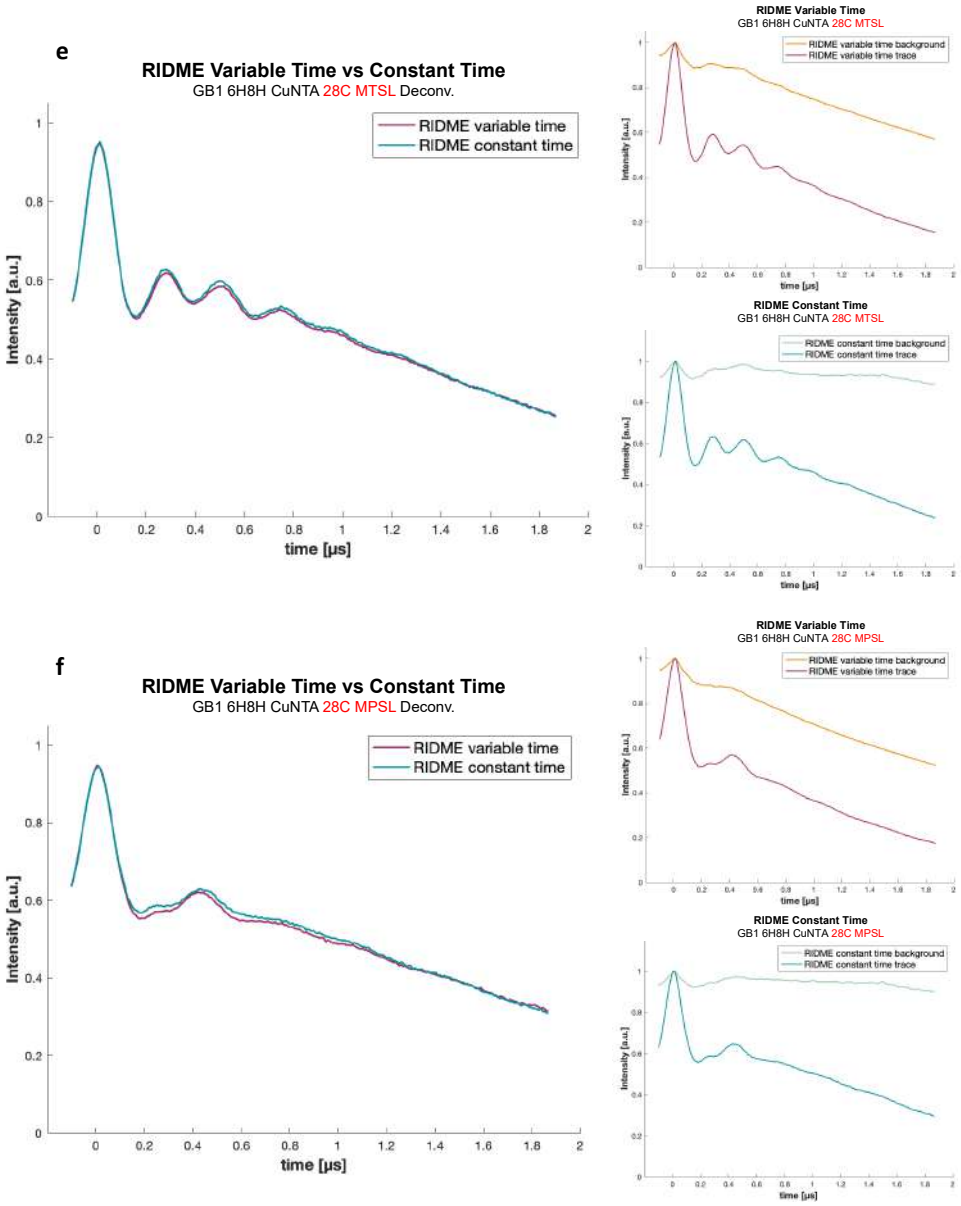


Fig. S3, continued Right: ctRIDME (cyan) and vtRIDME (magenta) raw data traces and their respective background traces (light green and orange). Left: superimposed constant time and variable time RIDME traces after deconvolution **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

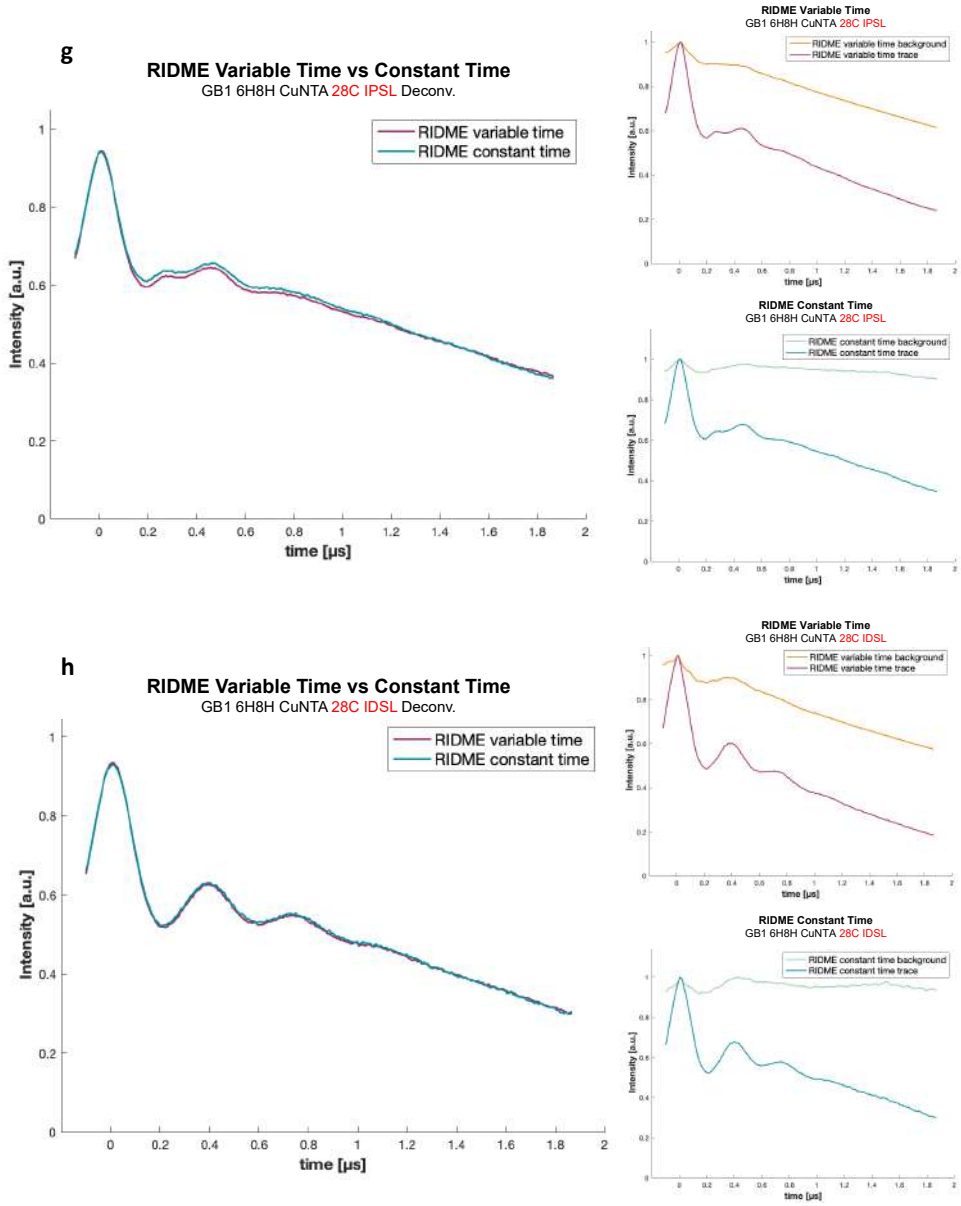


Fig. S3, continued Right: ctRIDME (cyan) and vtRIDME (magenta) raw data traces and their respective background traces (light green and orange). Left: superimposed constant time and variable time RIDME traces after deconvolution **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

1.4 RIDME processed spectra and distance distributions

Raw and background-corrected RIDME traces, processed with Deer Analysis, and their corresponding distance distributions are reported in **Fig. S4**.

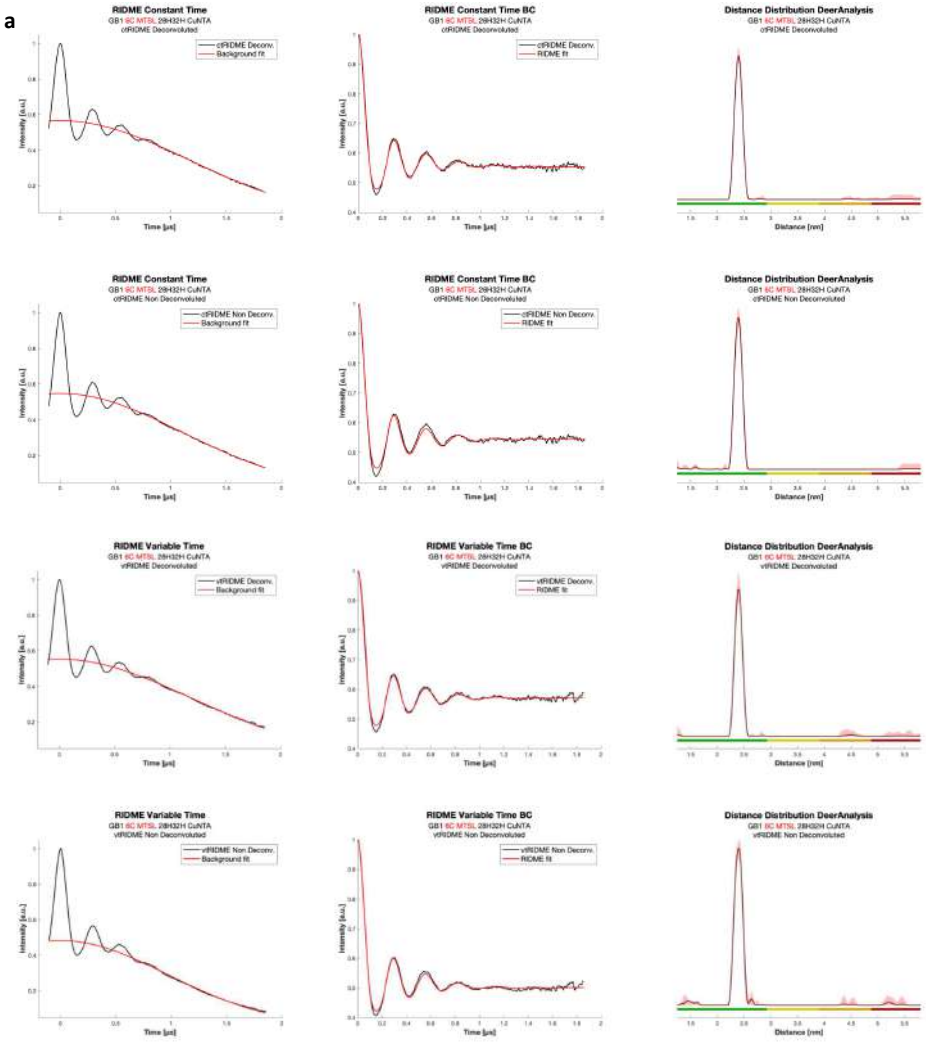


Fig. S4 Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted. **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

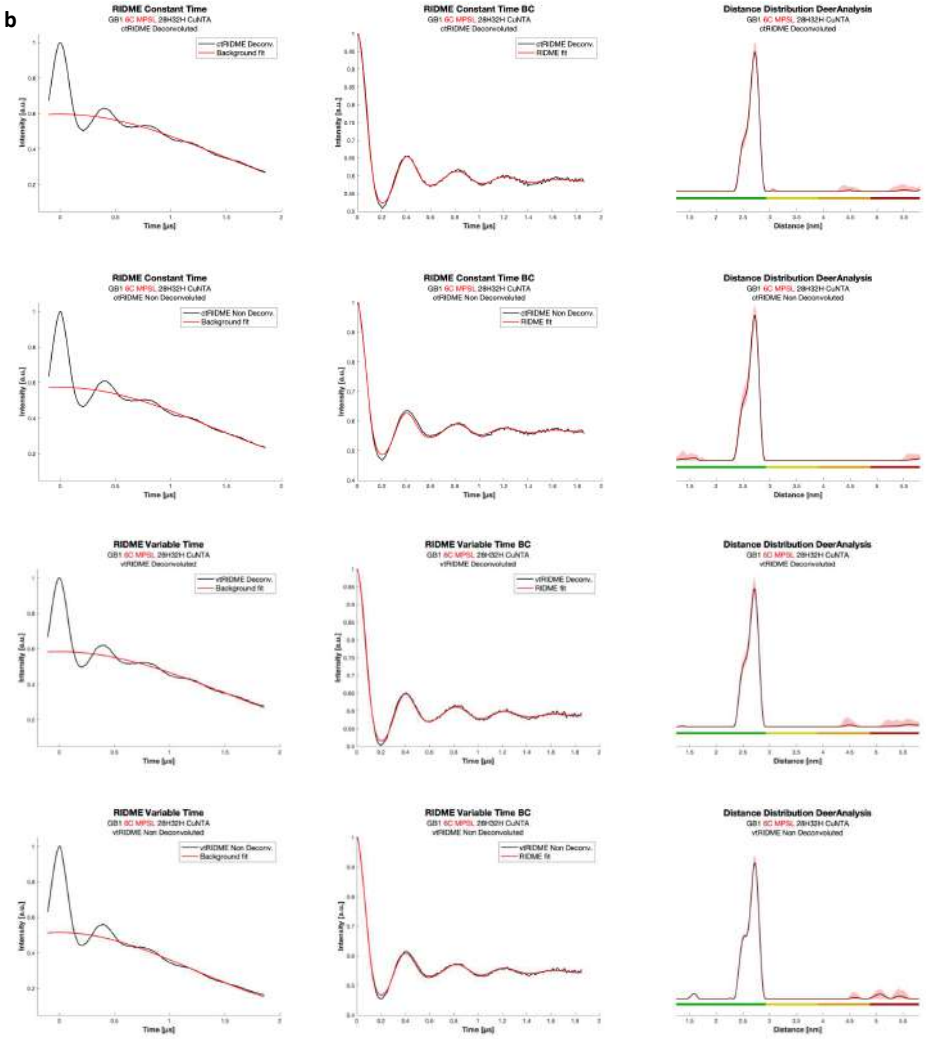


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

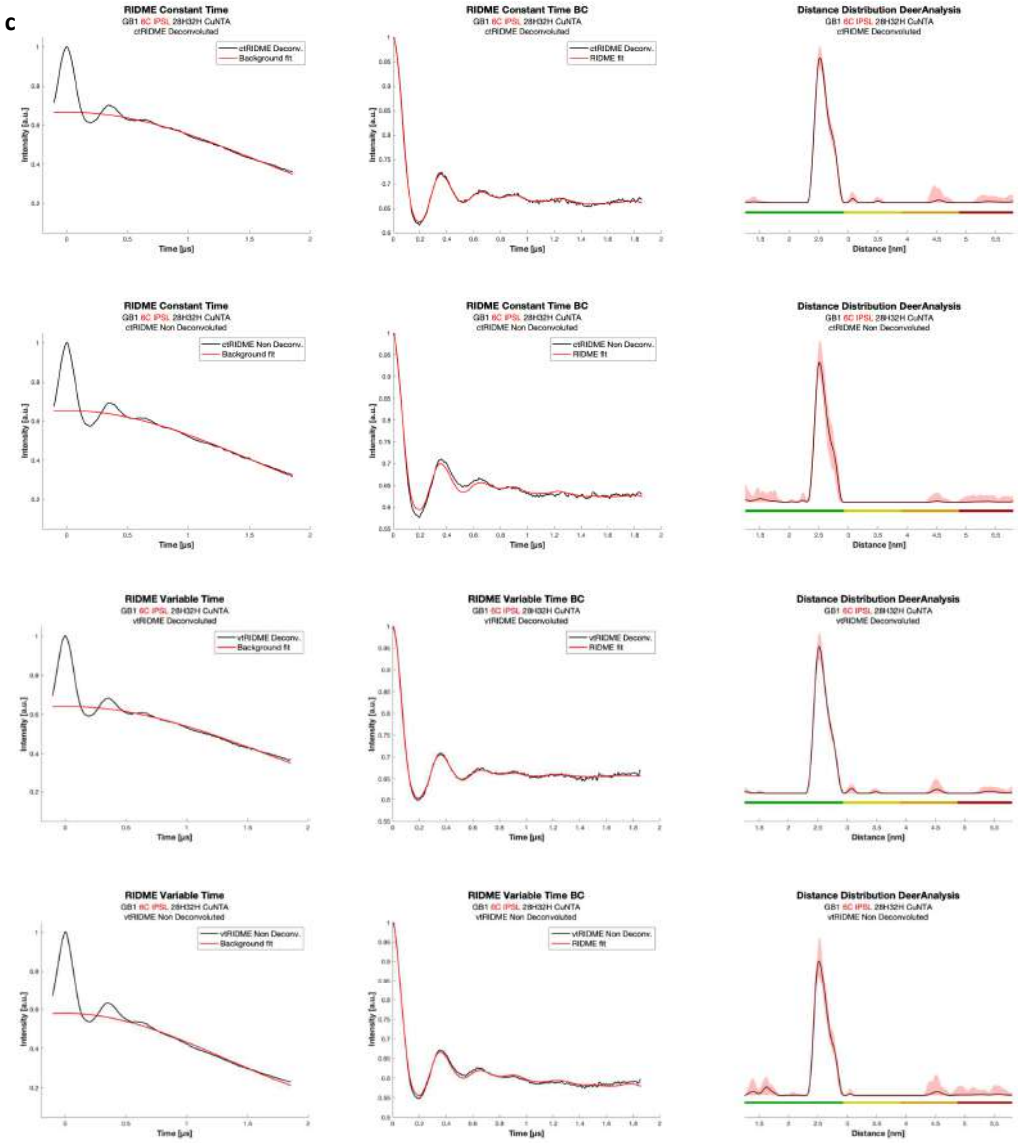


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

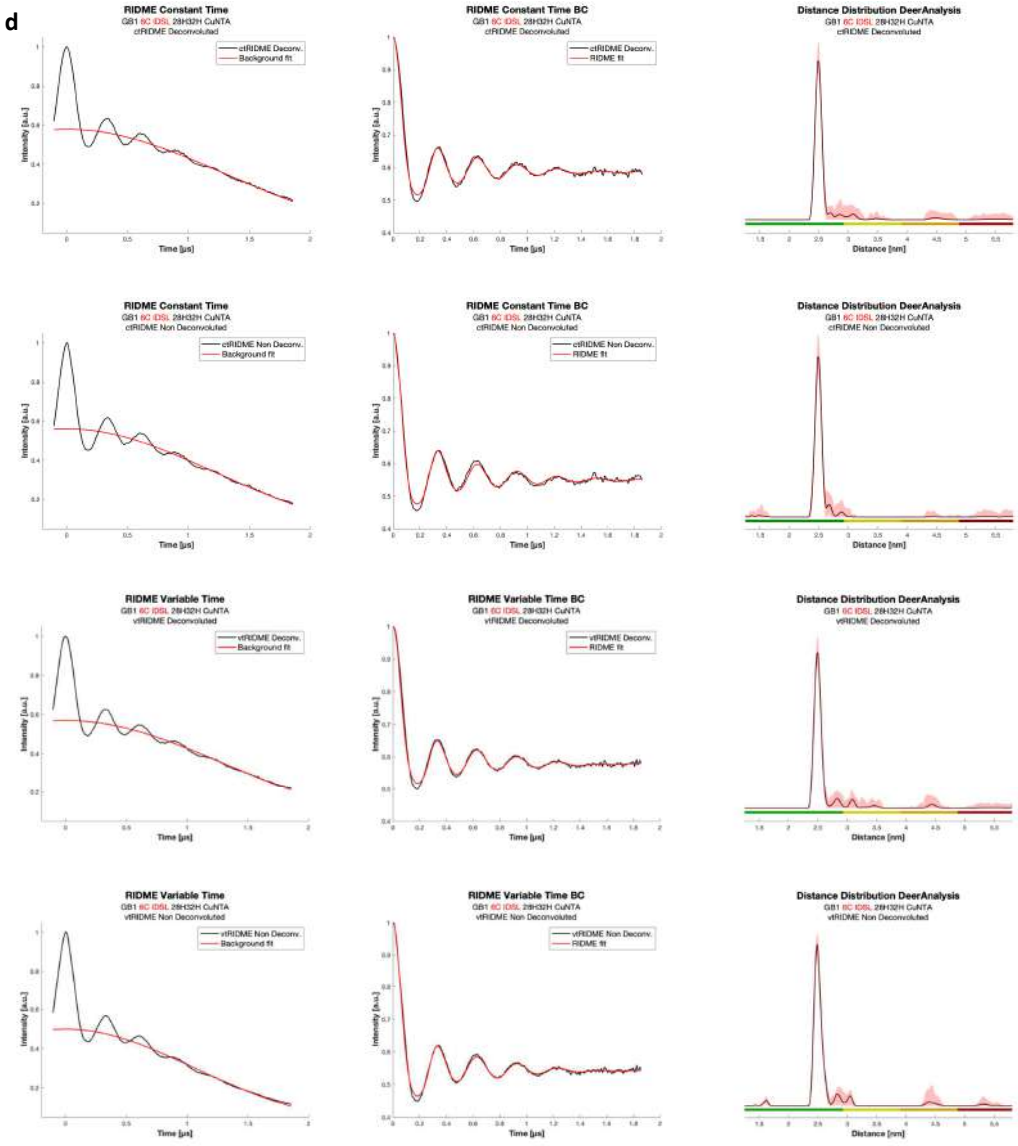


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

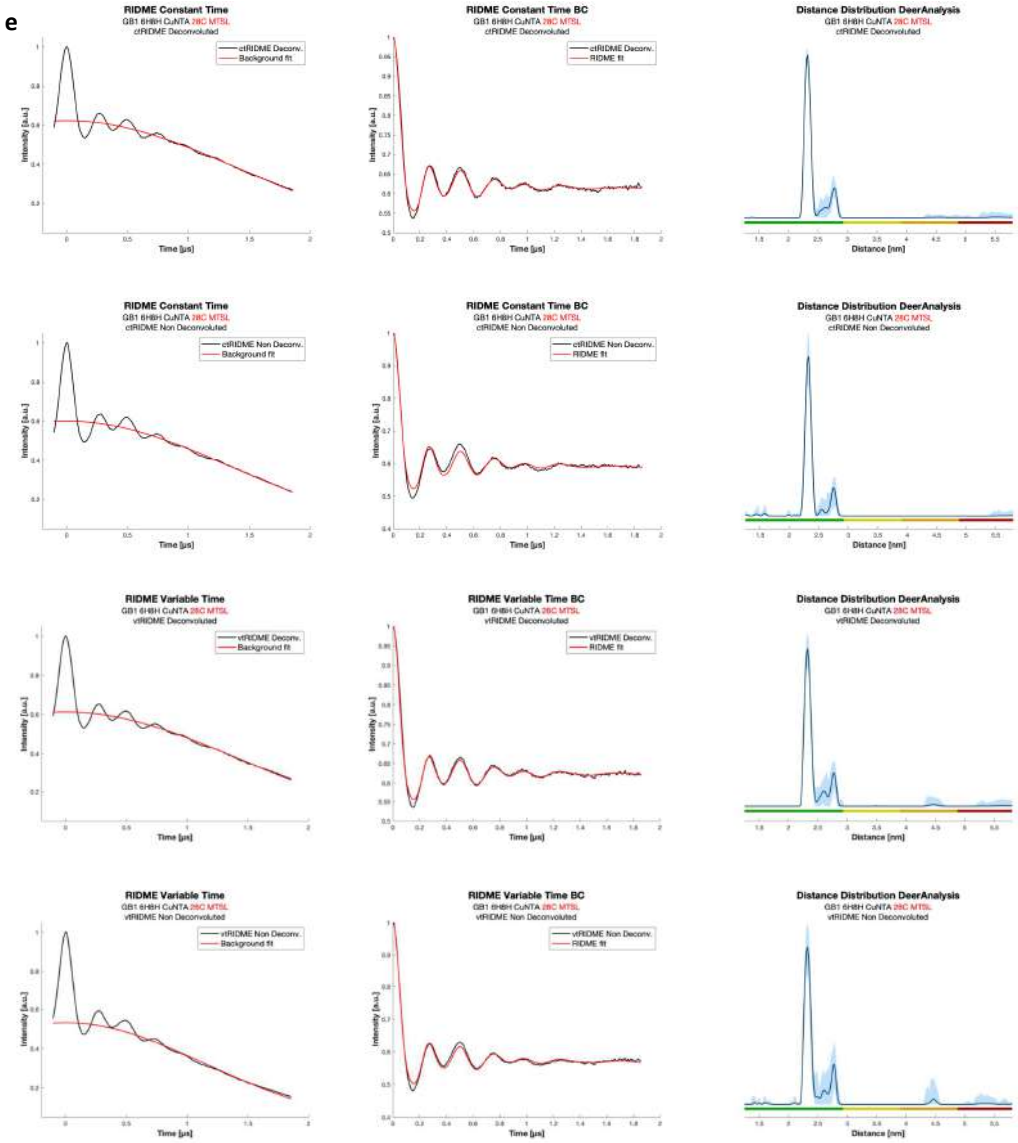


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

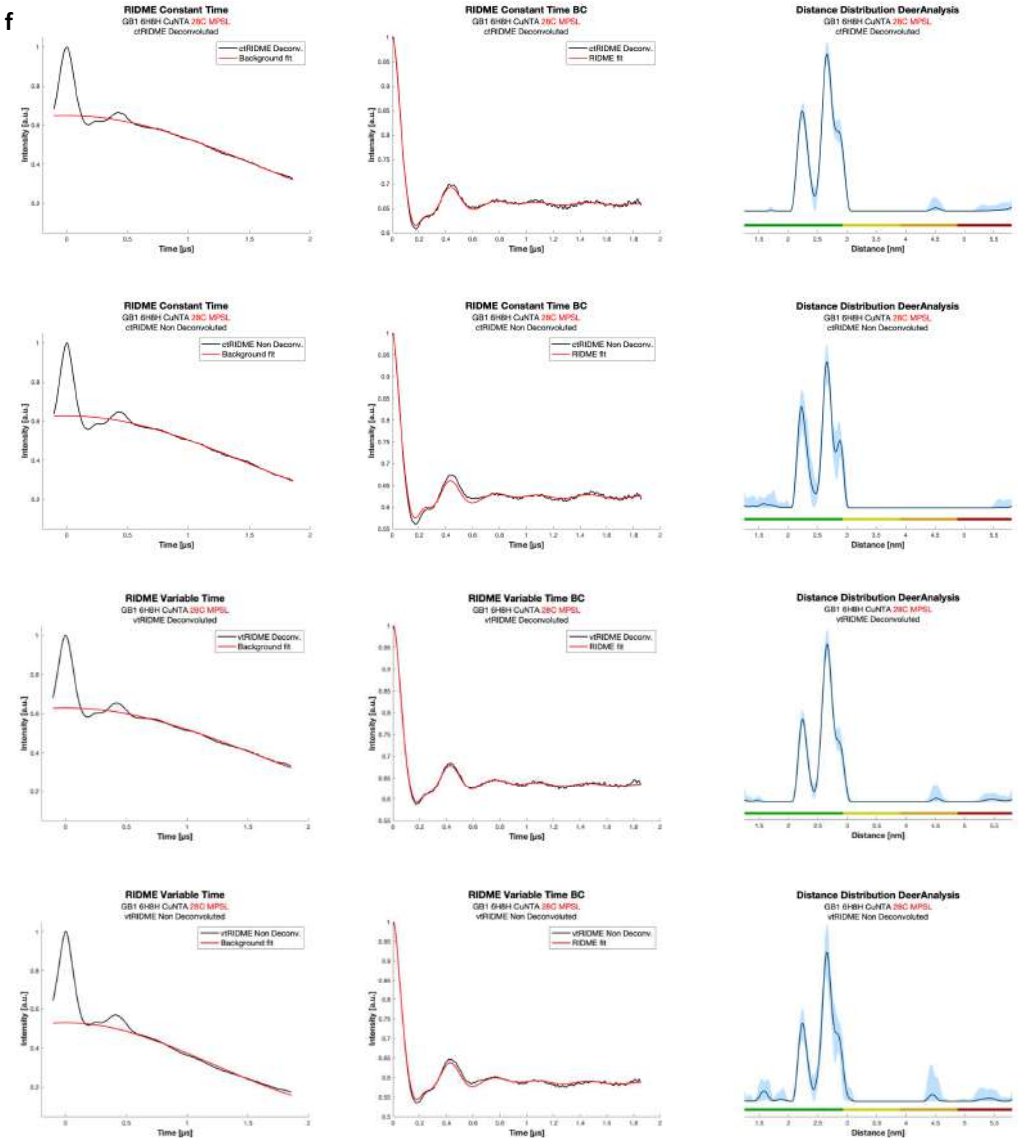


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

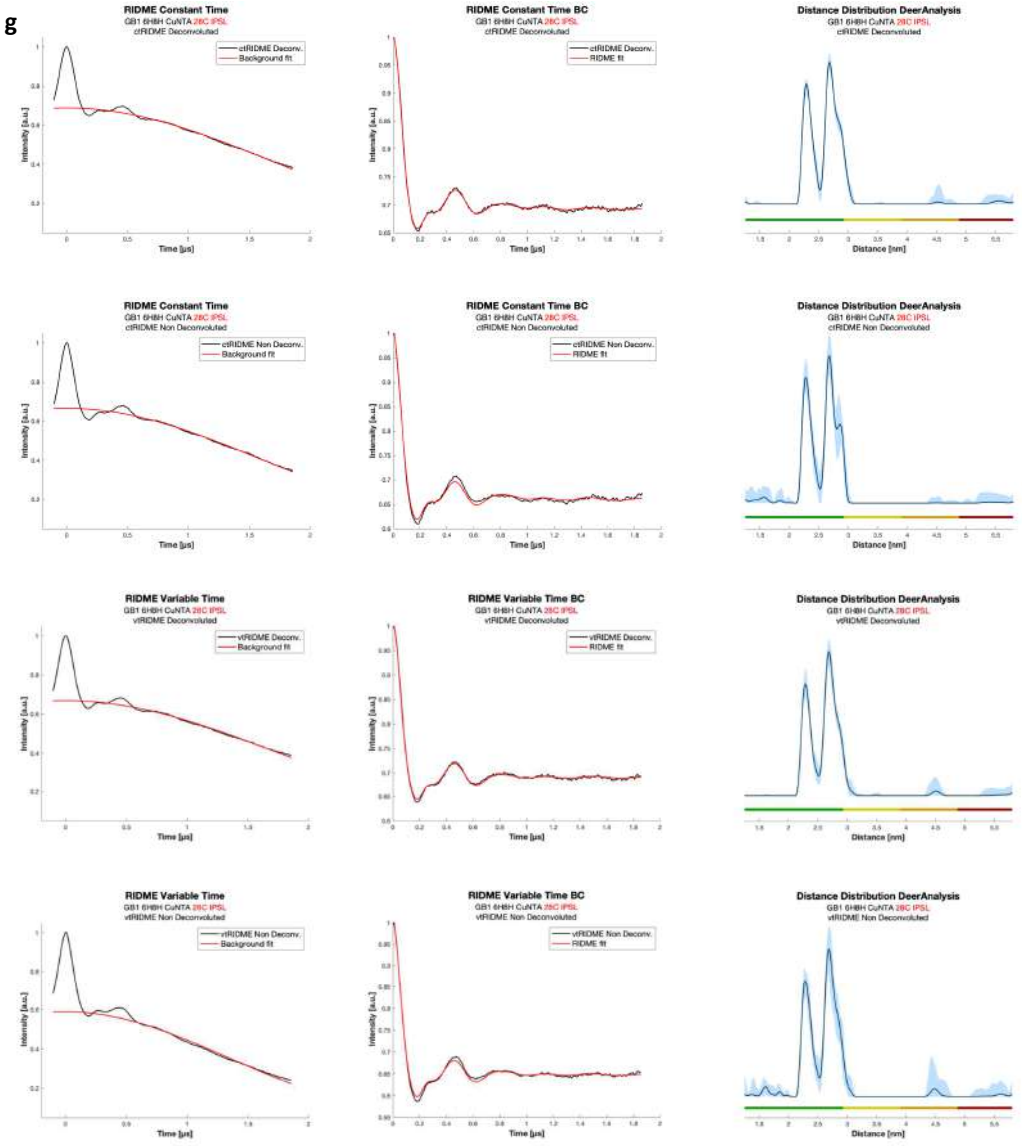


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

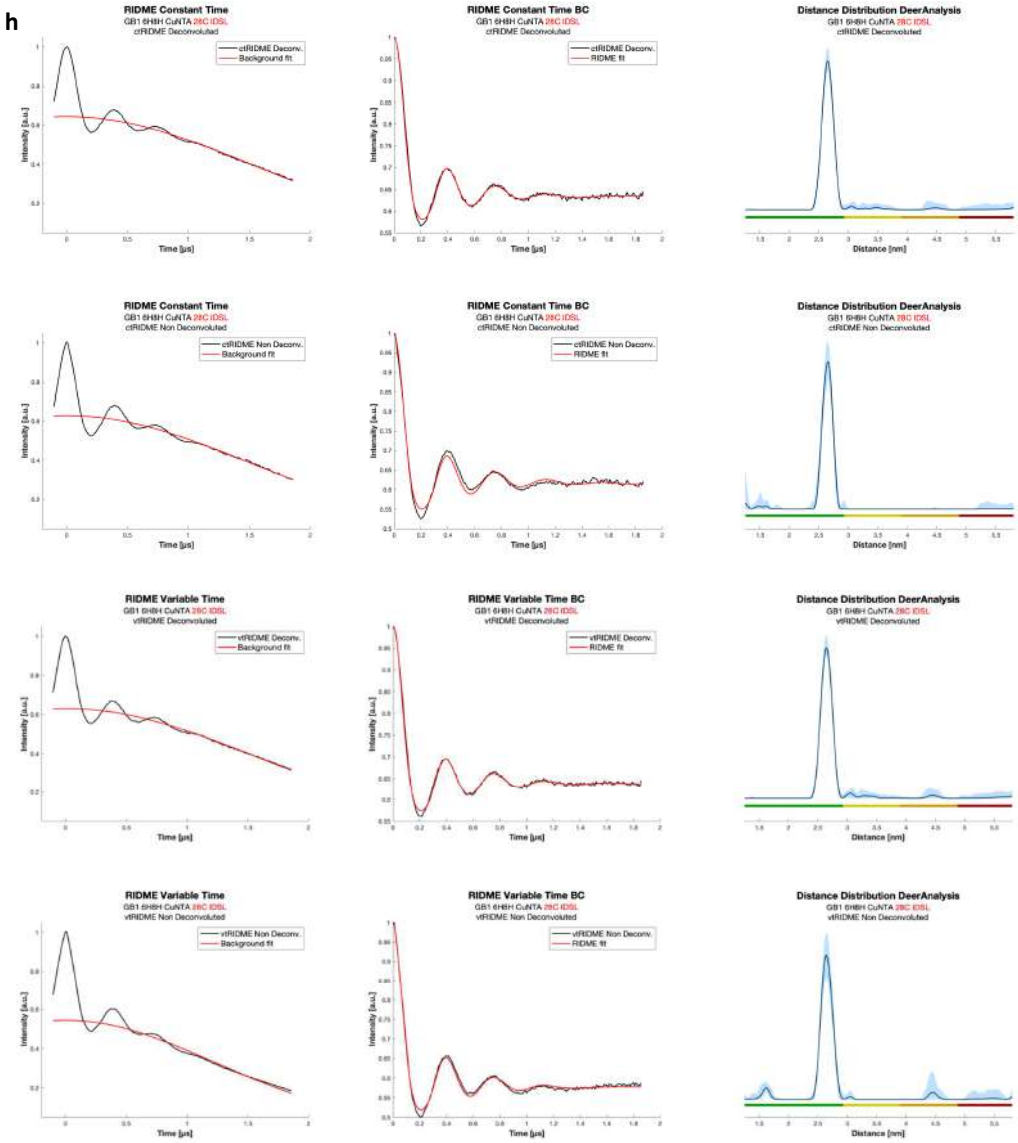


Fig. S4, continued Left: raw RIDME traces (black) with background function (red); middle: background-corrected (BC) data (black) with fit (red); right: corresponding distance distributions given as 95% confidence intervals (± 2σ) with 50% noise added for error estimation during statistical analysis. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). First row: ctRIDME deconvoluted, second row: ctRIDME non deconvoluted, third row: vtRIDME deconvoluted, fourth row: vtRIDME non deconvoluted **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

1.5 Sensitivity data for vtRIDME and ctRIDME

We employed both ctRIDME and vtRIDME to compare the two pulse sequences in their performances for gaining confidence in the general applicability and robustness of the variable-time sequence. Therefore, the modulation depth extracted from each RIDME trace during processing in DeerAnalysis, divided by the noise (RMSD), calculated from the phase corrected imaginary part of the data (**Table S3**), were employed for obtaining the sensitivity per echo and the sensitivity per unit of time for each trace (**Table S4**) as reported previously[1]. As expected, higher sensitivity values were achieved from the vtRIDME, especially when considering the ratio between the non deconvoluted sequences.

ctRIDME vs vtRIDME

Sensitivities		Δ				RMSD				S			
		ctRIDME		vtRIDME		ctRIDME		vtRIDME		ctRIDME		vtRIDME	
		Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.
MTSL	6C	0.434	0.452	0.447	0.517	0.0030	0.0030	0.0024	0.0016	142.8	149.9	187.6	323.1
	28C	0.377	0.406	0.387	0.428	0.0023	0.0026	0.0020	0.0014	193.8	189.8	232.5	313.5
MPSL	6C	0.403	0.427	0.416	0.486	0.0021	0.0022	0.0018	0.0015	159.4	134.1	166.3	237.7
	28C	0.351	0.373	0.370	0.470	0.0020	0.0022	0.0018	0.0011	151.1	151.0	222.7	351.1
IPSL	6C	0.333	0.346	0.358	0.420	0.0021	0.0026	0.0021	0.0018	164.5	157.3	196.0	304.2
	28C	0.312	0.334	0.331	0.410	0.0017	0.0024	0.0015	0.0011	171.4	166.4	207.0	416.2
IDSL	6C	0.412	0.440	0.430	0.501	0.0027	0.0029	0.0019	0.0014	185.5	138.4	226.2	385.5
	28C	0.357	0.372	0.369	0.419	0.0024	0.0028	0.0025	0.0019	147.4	133.0	146.4	219.1

Table S3 Noise estimates (RMSD), modulation depths (Δ) and the sensitivity obtained from their ratio (S) for the two different RIDME pulse sequences (variable or constant time) with and without deconvolution for the two GB1 constructs (I6C/K28H/Q32H and I6H/N8H/K28C) with the four nitroxide labels, MTSL, MPSL, IPSL, IDSL

ctRIDME vs vtRIDME

Sensitivities		S_n				S_t				S_{ratio}	
		ctRIDME		vtRIDME		ctRIDME		vtRIDME			
		Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	vt/ct Deconvoluted	vt/ct Non Deconvoluted
MTSL	6C	4.464	4.686	5.864	10.09	44.64	46.86	58.64	100.9	1.314	2.155
	28C	6.058	5.933	7.267	9,797	49.82	59.33	72.67	97.97	1.191	1.934
MPSL	6C	4.982	4.192	5.199	7.431	49.82	41.92	51.99	74.31	1.119	1.651
	28C	4.725	4,720	6,961	10,97	47.25	47.20	69.61	109.7	1.208	2.501
IPSL	6C	5.141	4.917	6.126	9.509	51.41	49.17	61.26	95.09	1.044	1.772
	28C	5.359	5.201	6.471	13.00	53.59	52.01	64.71	130.0	1.219	2.784
IDSL	6C	5.800	4.328	7.070	12.05	58.00	43.28	70.70	120.5	1.473	2.325
	28C	4.607	4.158	4.575	6.848	46.07	41.58	45.75	68.48	0.933	1.647

Table S4 Normalized sensitivity values (S_n) obtained from the sensitivity (S, **Table S3**) divided by the square root of total echoes per point (taken as the product of number of scans (1), shots per point (2), number of τ averages (16), and phase cycle(32)), and sensitivity per unit of time (S_t) obtained by multiplying S_n values for the square root of the averaging rate (100 Hz). The values were extracted for both the constant and variable time RIDME with and without deconvolution for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C with the four nitroxide labels MTSL, MPSL, IPSL and IDSL. S_{ratio} represents the ratio between the sensitivity values for the constant and variable time RIDME, respectively for the deconvoluted or non deconvoluted.

1.6 Distance distributions extracted from vtRIDME and ctRIDME with and without deconvolution

As expected from systems with short distances and relatively narrow distance distributions, the two different RIDME pulse sequences do not display significant discrepancies, neither in shape nor in mean and widths (**Table S4**), showing high consistency and robustness, regardless whether the deconvolution step has been applied or not. However, the distance distributions belonging to the vtRIDME without the deconvolution step (in green) are the most affected by the presence of artifacts at higher distance ranges. The superimposed distance distributions for both GB1 constructs, each one with the four distinct nitroxide labels, are reported in **Fig. S5**.

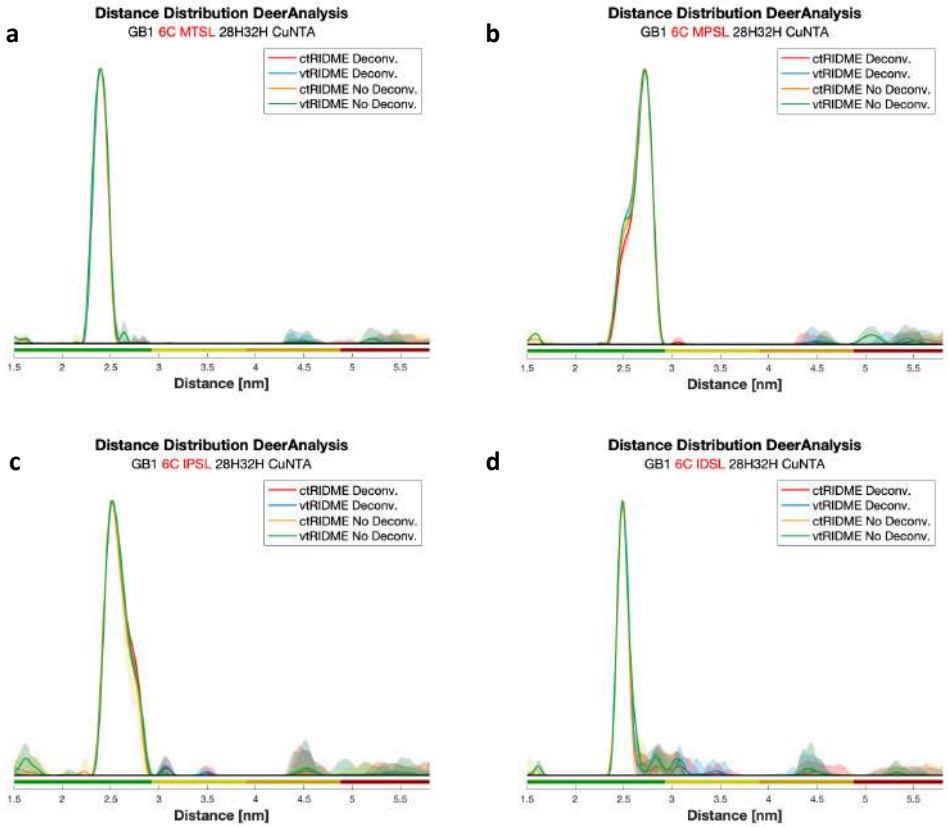


Fig. S5 Superimposed distance distributions extracted from ctRIDME and vtRIDME both deconvoluted and non deconvoluted traces, with the confidence estimates intervals (± 2σ) shown as shaded regions. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

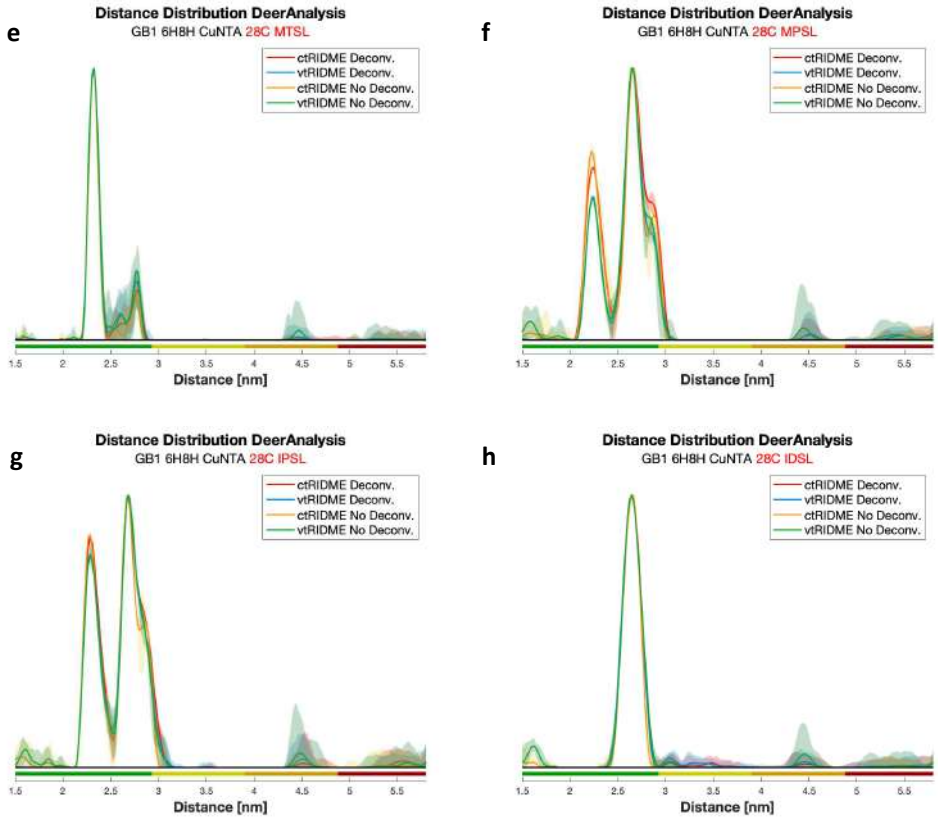


Fig. S5, continued Superimposed distance distributions extracted from ctRIDME and vtRIDME both deconvoluted and non deconvoluted traces, with the confidence estimates intervals (± 2σ) shown as shaded regions. Colour bars represent reliability ranges (green: shape reliable; yellow: mean and width reliable; orange: mean reliable: red: no quantification possible). **a)** GB1 I6C/K28H/Q32H MTSL CuNTA, **b)** GB1 I6C/K28H/Q32H MPSL CuNTA, **c)** GB1 I6C/K28H/Q32H IPSL CuNTA, **d)** GB1 I6C/K28H/Q32H IDSL CuNTA, **e)** GB1 I6H/N8H/K28C CuNTA MTSL, **f)** GB1 I6H/N8H/K28C CuNTA MPSL, **g)** GB1 I6H/N8H/K28C CuNTA IPSL, **h)** GB1 I6H/N8H/K28C CuNTA IDSL

1.7 Mean, Width and Full Width at Half Maximum for ctRIDME and vtRIDME both with and without deconvolution

All mean and width values of the experimental distance distributions were extracted with an in-house Matlab script. The values were calculated considering only the data between 1.7 and 4.5 nm to suppress the influence of artifacts at higher or lower distance ranges that greatly affect these values (**Table S5**). Mean values for the different experimental setups on the same sample can be considered consistent and robust with respect to each other. On the other hand, the vtRIDME suffers the most from the presence of artifacts which is reflected in the consistently higher values of the extracted widths.

ctRIDME vs vtRIDME		Mean (nm)				Width (nm)			
		ctRIDME		vtRIDME		ctRIDME		vtRIDME	
		Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.
MTSL	6C	2.402	2.390	2.412	2.407	0.131	0.062	0.189	0.159
	28C	2.404	2.382	2.424	2.453	0.185	0.154	0.244	0.341
MPSL	6C	2.674	2.653	2.662	2.657	0.168	0.115	0.171	0.124
	28C	2.581	2.552	2.591	2.611	0.272	0.245	0.261	0.324
IPSL	6C	2.596	2.574	2.601	2.601	0.185	0.149	0.208	0.212
	28C	2.592	2.565	2.600	2.607	0.256	0.233	0.277	0.317
IDSL	6C	2.561	2.515	2.596	2.589	0.259	0.133	0.333	0.319
	28C	2.682	2.641	2.692	2.705	0.217	0.091	0.250	0.308

Table S5 Mean and width values (reported in nm) extracted from experimental distance distributions of the two RIDME pulse sequences with and without deconvolution for both GB1 constructs (I6C/K28H/Q32H and I6H/N8H/K28C) for the four spin labels MTSL, MPSL, IPSL, IDSL

However, in some cases considering the restricted range between 1.7 and 4.5 nm is not enough to be completely independent from artifacts, and some values are still relatively affected by them, such as the width for the 28C IDSL. Therefore, we monitored the full width at half maximum (FWHM), considering this to be a parameter more independent from artifacts, for every distance distribution peak, to further point out the high consistency and robustness of the two different RIDME sequences (**Table S6**). Moreover, we decided to extract the mean and width values for every distance distribution considering the lower error estimate obtained from Deer Analysis (**Table S7**).

ctRIDME vs vtRIDME

FWHM (nm)		MTSL		MPSL		IPSL		IDSL	
		6C	28C	6C	28C	6C	28C	6C	28C
ctRIDME	Deconv.	0.179	0.125	0.221	0.688	0.285	0.669	0.133	0.229
	No Deconv.	0.180	0.120	0.227	0.590	0.256	0.672	0.128	0.212
vtRIDME	Deconv.	0.175	0.128	0.246	0.557	0.274	0.635	0.139	0.230
	No Deconv.	0.191	0.130	0.220	0.554	0.265	0.640	0.139	0.231

Table S6 Full width at half maximum (FWHM) (reported in nm) extracted from experimental distance distributions of the two RIDME pulse sequences with and without deconvolution for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C for the four spin labels MTSL, MPSL, IPSL, IDSL

ctRIDME vs vtRIDME		Mean LB (nm)				Width LB (nm)			
		ctRIDME		vtRIDME		ctRIDME		vtRIDME	
		Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.	Deconv.	Non Deconv.
MTSL	6C	2.393	2.390	2.395	2.391	0.063	0.060	0.062	0.063
	28C	2.361	2.340	2.350	2.339	0.135	0.102	0.113	0.108
MPSL	6C	2.669	2.655	2.655	2.653	0.107	0.111	0.111	0.114
	28C	2.576	2.557	2.582	2.561	0.241	0.235	0.219	0.219
IPSL	6C	2.578	2.556	2.577	2.579	0.114	0.100	0.115	0.113
	28C	2.581	2.553	2.579	2.566	0.233	0.228	0.226	0.222
IDSL	6C	2.488	2.489	2.492	2.504	0.047	0.047	0.051	0.072
	28C	2.647	2.641	2.644	2.647	0.085	0.071	0.084	0.084

Table S7 Mean and width values (reported in nm) extracted from the lower error bound (LB) estimate of the experimental distance distributions of the two RIDME pulse sequences with and without deconvolution for both GB1 constructs (I6C/K28H/Q32H and I6H/N8H/K28C) for the four spin labels MTSL, MPS, IPSL and IDSL

1.8 Comparison of the distance distributions between two nitroxides labels or between the copper(II) and the nitroxides

Data obtained previously on the double-cysteine GB1 construct (GB1 I6C/K28C) showed bimodality for the MTSL label and generally broader distributions for the MPSL and IPSL labels (**Fig. S6**). Distinguishing whether this behaviour depends on the nitroxide attached to the α-helix or the β-sheet was possible only after the introduction of the less conformationally flexible CuNTA chelator ligand alternatively on one labelling site of the protein. The reduction in the distribution widths, provoked by the rigidity of this bipedal ligand, entailed a significant improvement in the precision of the measured distances, and a consequent decrease in the ambiguity of the interpretation of the system behaviour, leading us to mark the helix site as the one responsible for the broadness and bimodality in the distance distributions.

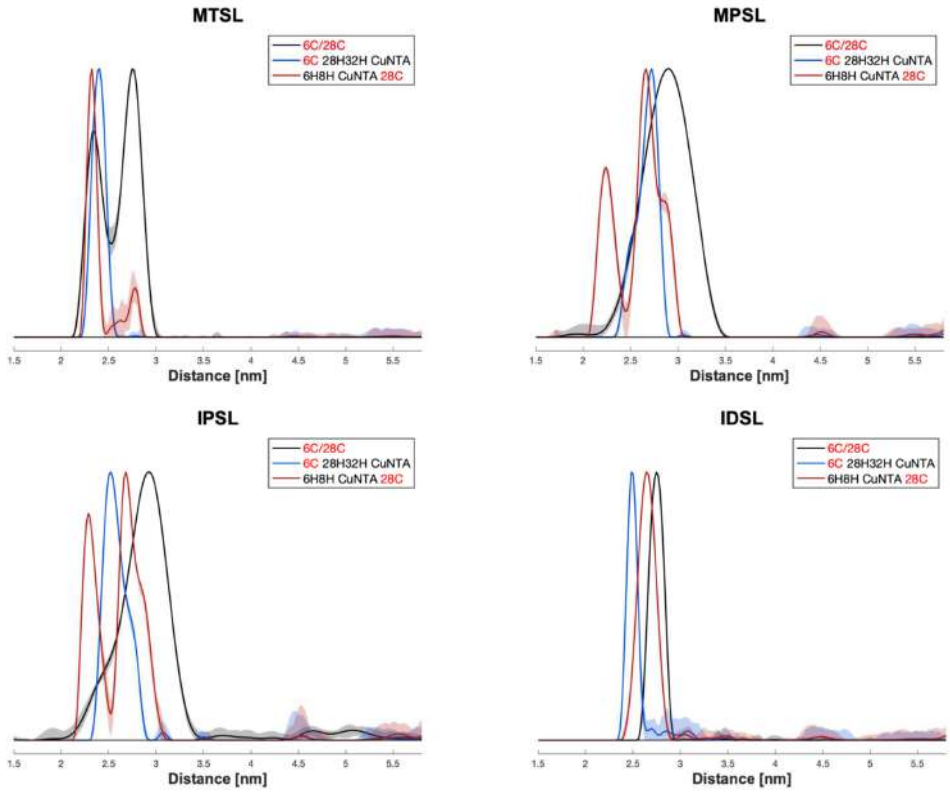


Fig. S6 Comparison of the distance distributions between two nitroxide spin labels on the double cysteine mutant GB1 I6C/K28C (black) obtained with the 4-pulse DEER sequence, and the distributions between the CuNTA and the same nitroxides, respectively on the I6C/K28H/Q32H (blue) and I6H/N8H/K28C (red) GB1 constructs obtained with the 5-pulse RIDME pulse sequence. The shadowed area represents the confidence estimate intervals (± 2σ)

1.9 Modelling with MMM and MtsslWizard

All the modelling data of the distance distributions, acquired for the different structure prediction tools (AlphaFold2 (**Fig. S7**), OmegaFold (**Fig. S8**), ESMFold (**Fig. S9**)), and the crystallographic structure (PDB:4wh4 (**Fig. S10**)), are reported here compared to the experimentally obtained distributions from the ctRIDME deconvoluted data set. All structural models behave in a similar manner, and the extensive discussion in the main manuscript about the AlphaFold2 performance can be transposed to the other prediction methods and to the crystallographic structure. Minor differences in terms of shapes of the distance distributions can be detected. Here we report a visual representation of the computated rotamers for the different labelling approaches and their respective different conditions (ambient and cryogenic temperature for MMM and Tight and Loose for MtsslWizard) for the nitroxides (**Fig. S11**) and for the CuNTA (Tight and Loose settings for the Wizard and as previously described for MMM [2]) (**Fig. S12**).

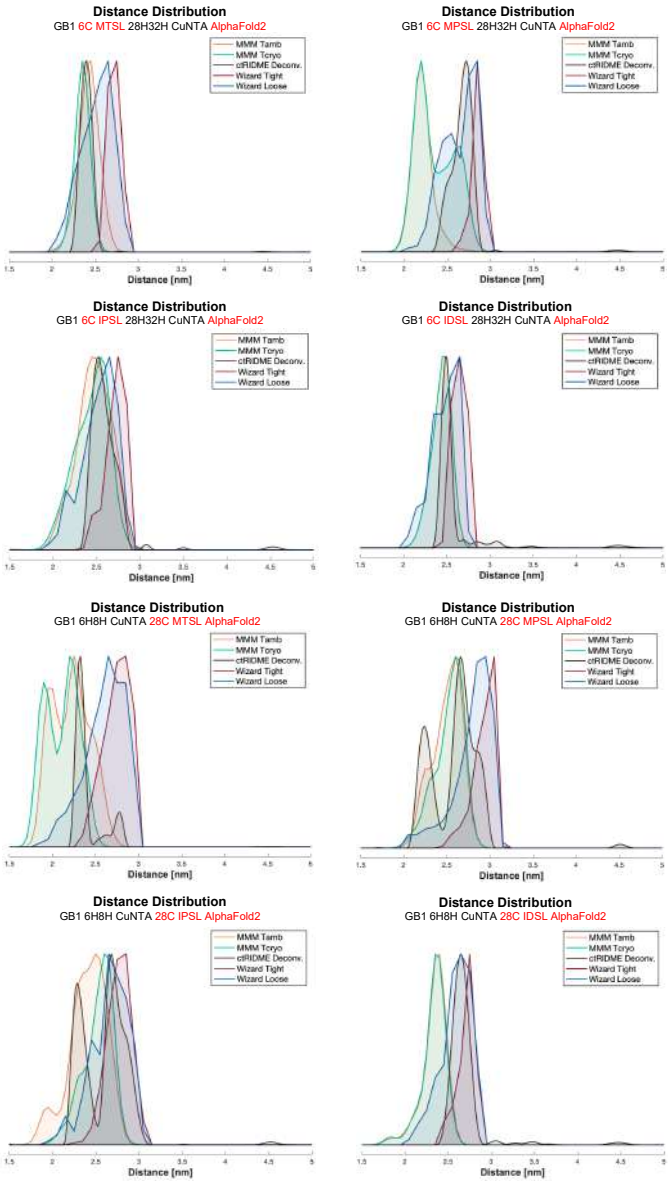


Fig. S7 Modelled distance distributions for both I6C/K28H/Q32H and the I6H/N8H/K28C GB1 constructs, both labelled with MTSL, MPSL, IPSL and IDSL, based on the AlphaFold2 structure, superimposed with their respective experimental distance distributions (derived from ctRIDME deconvoluted data set, in black). The *in silico* approaches considered are MMM at ambient (orange) and cryogenic (green) temperature and MtsslWizard with Tight (red) and Loose (blue) settings. First row: 6C MTSL, 6C MPSL. Second row: 6C IPSL, 6C IDSL. Third row: 28C MTSL and 28C MPSL. Fourth row: 28C IPSL and 28C IDSL

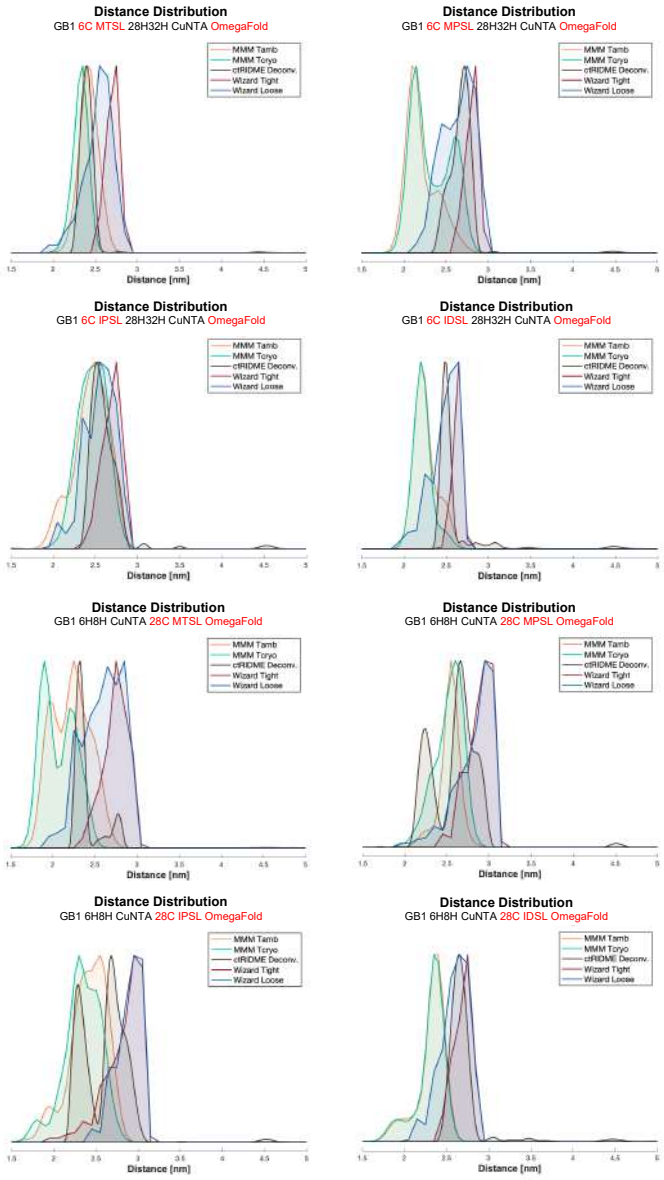


Fig. S8 Modelled distance distributions for both I6C/K28H/Q32H and the I6H/N8H/K28C GB1 constructs, both labelled with MTSL, MPSL, IPSL and IDSL, based on the OmegaFold structure, superimposed with their respective experimental distance distributions (derived from ctRIDME deconvoluted data set, in black). The *in silico* approaches considered are MMM at ambient (orange) and cryogenic (green) temperature and MtsslWizard with Tight (red) and Loose (blue) settings. First row: 6C MTSL, 6C MPSL. Second row: 6C IPSL, 6C IDSL. Third row: 28C MTSL and 28C MPSL. Fourth row: 28C IPSL and 28C IDSL

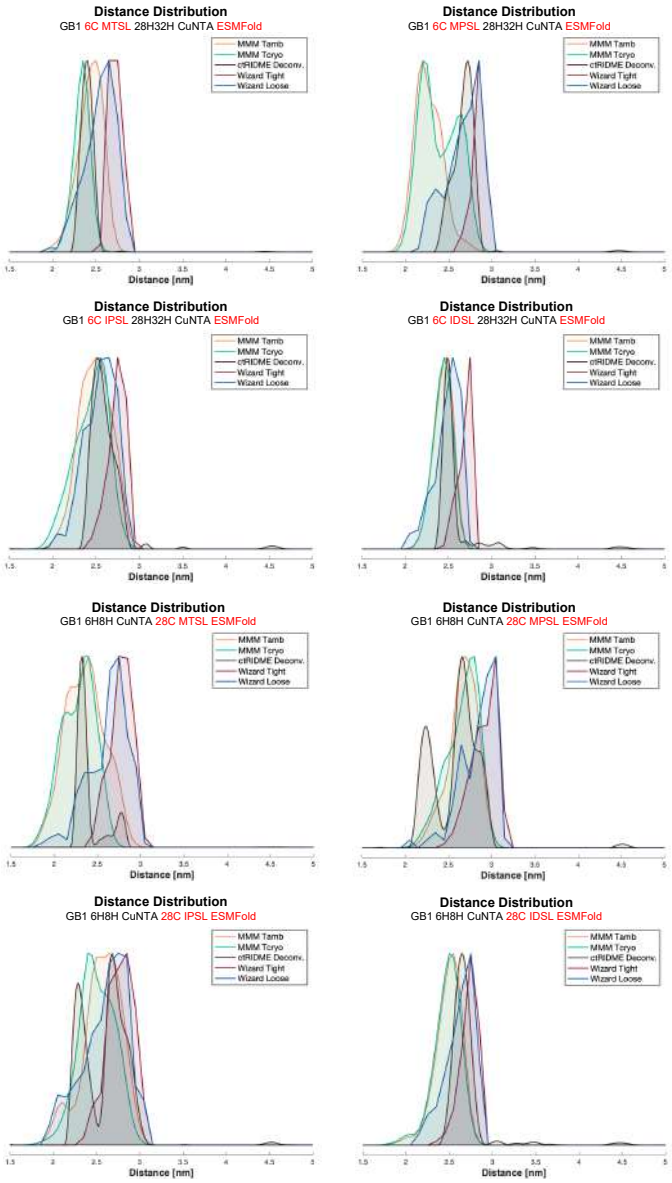


Fig. S9 Modelled distance distributions for both I6C/K28H/Q32H and the I6H/N8H/K28C GB1 constructs, both labelled with MTSL, MPSL, IPSL and IDSL, based on the ESMFold structure, superimposed with their respective experimental distance distributions (derived from ctRIDME deconvoluted data set, in black). The *in silico* approaches considered are MMM at ambient (orange) and cryogenic (green) temperature and MtsslWizard with Tight (red) and Loose (blue) settings. First row: 6C MTSL, 6C MPSL. Second row: 6C IPSL, 6C IDSL. Third row: 28C MTSL and 28C MPSL. Fourth row: 28C IPSL and 28C IDSL

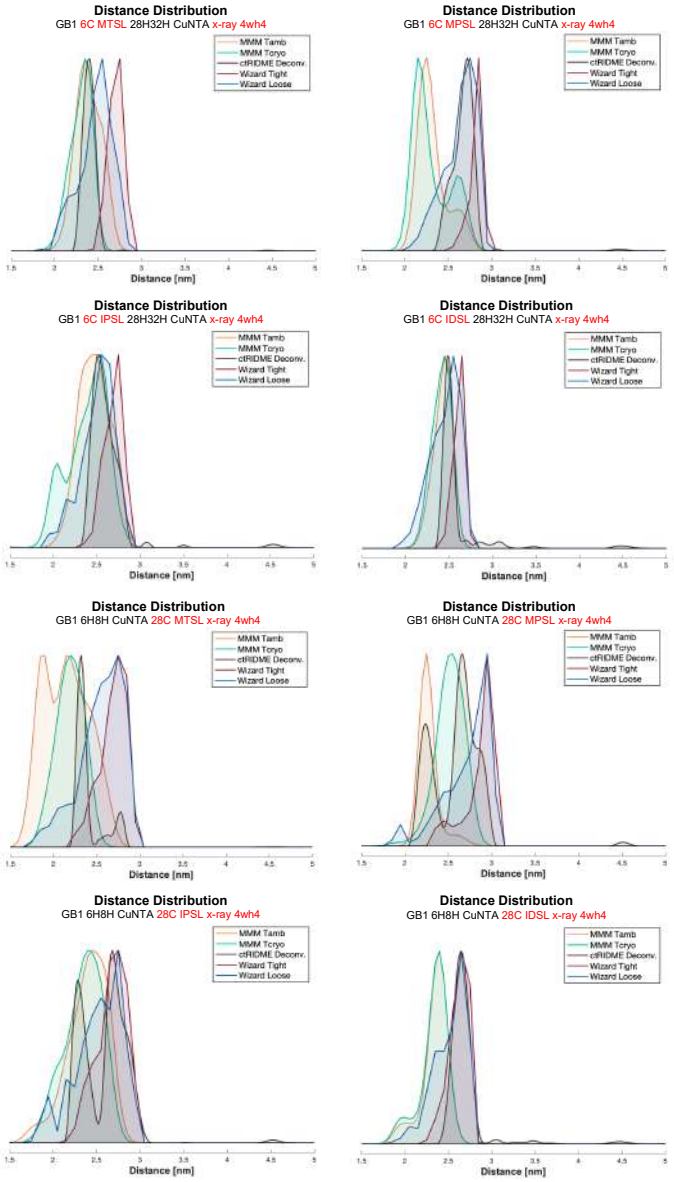


Fig. S10 Modelled distance distributions for both I6C/K28H/Q32H and the I6H/N8H/K28C GB1 constructs, both labelled with MTSL, MPSL, IPSL and IDSL, based on the X-ray structure (PDB: 4wh4), superimposed with their respective experimental distance distributions (derived from ctRIDME deconvoluted data set, in black). The *in silico* approaches considered are MMM at ambient (orange) and cryogenic (green) temperature and MtsslWizard with Tight (red) and Loose (blue) settings. First row: 6C MTSL, 6C MPSL. Second row: 6C IPSL, 6C IDSL. Third row: 28C MTSL and 28C MPSL. Fourth row: 28C IPSL and 28C IDSL

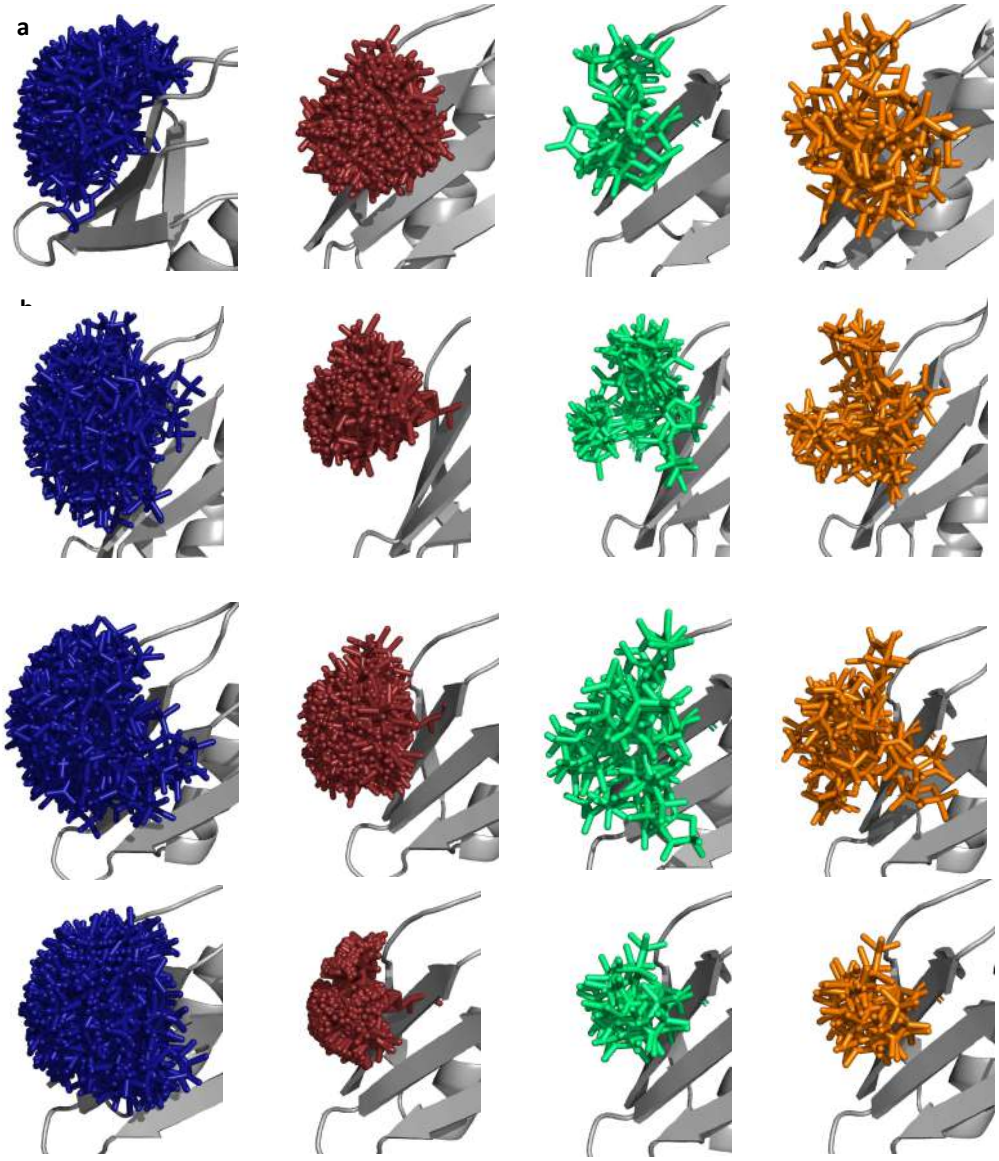


Fig. S11 Modelled rotamers for the two in silico labelling approaches, MMM with cryogenic (green rotamers) and ambient (orange rotamers) conditions, and MtsslWizard with Tight (red rotamers) and Loose (blue rotamers) settings, for both I6C/K28H/Q32H and the I6H/N8H/K28C GB1 constructs, labelled with MTSL, MPSL, IPSL and IDSL. **a)** GB1 I6C/K28H/Q32H MTSL, **b)** GB1 I6C/K28H/Q32H MPSL, **c)** GB1 I6C/K28H/Q32H IPSL, **d)** GB1 I6C/K28H/Q32H IDSL, **e)** GB1 I6H/N8H/K28C MTSL, **f)** GB1 I6H/N8H/K28C MPSL, **g)** GB1 I6H/N8H/K28C IPSL, **h)** GB1 I6H/N8H/K28C IDSL

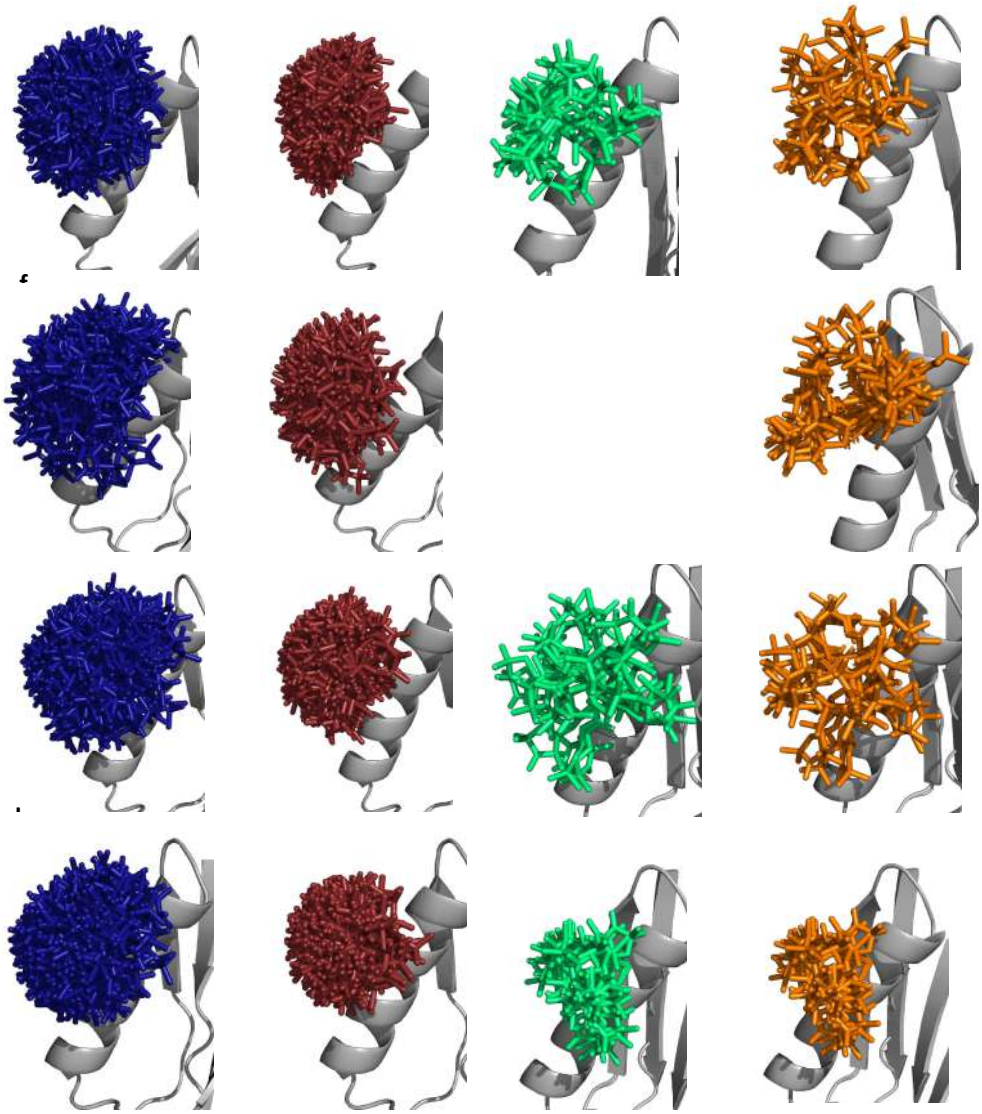


Fig. S11, continued Modelled rotamers for the two in silico labelling approaches, MMM with cryogenic (green rotamers) and ambient (orange rotamers) conditions, and MtsslWizard with Tight (red rotamers) and Loose (blue rotamers) settings, for both I6C/K28H/Q32H and the I6H/N8H/K28C GB1 constructs, labelled with MTSL, MPSL, IPSL and IDSL. **a)** GB1 I6C/K28H/Q32H MTSL, **b)** GB1 I6C/K28H/Q32H MPSL, **c)** GB1 I6C/K28H/Q32H IPSL, **d)** GB1 I6C/K28H/Q32H IDSL, **e)** GB1 I6H/N8H/K28C MTSL, **f)** GB1 I6H/N8H/K28C MPSL, **g)** GB1 I6H/N8H/K28C IPSL, **h)** GB1 I6H/N8H/K28C IDSL

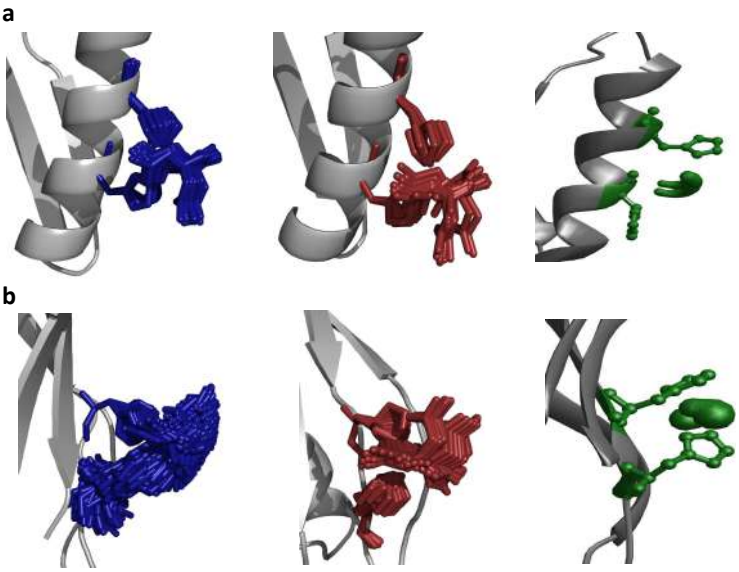


Fig. S12 Modelled rotamers for the CuNTA label for the two *in silico* labelling approaches MtsslWizard Tight (red rotamers) and Loose (blue rotamers) settings and for MMM (forest green rotamers) for both I6C/K28H/Q32H (**a**) and I6H/N8H/K28C (**b**) GB1 constructs

OmegaFold		Mean				Width				Δ Mean			
		MMM		MtsslWizard		MMM		MtsslWizard		MMM		MtsslWizard	
		T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose
MTSL	6C	2.323	2.424	2.704	2.534	0.091	0.114	0.077	0.175	0.078	0.022	0.301	0.132
	28C	2.071	2.283	2.746	2.594	0.200	0.194	0.146	0.234	0.332	0.121	0.341	0.189
MPSL	6C	2.333	2.212	2.805	2.638	0.237	0.188	0.075	0.193	0.340	0.462	0.130	0.036
	28C	2.521	2.533	2.888	2.824	0.171	0.125	0.149	0.222	0.060	0.048	0.306	0.242
IPSL	6C	2.466	2.451	2.690	2.547	0.171	0.208	0.115	0.187	0.130	0.145	0.093	0.049
	28C	2.342	2.410	2.753	2.609	0.219	0.223	0.145	0.239	0.249	0.181	0.161	0.017
IDSL	6C	2.223	2.257	2.631	2.479	0.115	0.138	0.051	0.167	0.337	0.303	0.070	0.081
	28C	2.308	2.317	2.690	2.595	0.177	0.175	0.109	0.168	0.373	0.365	0.079	0.086

Table S9 Mean, width and Δ mean of the modelled distances for both MMM (ambient and cryogenic temperature) and MtsslWizard (Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each labelled with the four nitroxides (MTSL, MPSL, IPSL and IDSL), based on the OmegaFold structure

ESMFold		Mean				Width				Δ Mean			
		MMM		MtsslWizard		MMM		MtsslWizard		MMM		MtsslWizard	
		T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose
MTSL	6C	2.334	2.434	2.723	2.547	0.096	0.147	0.075	0.178	0.067	0.032	0.320	0.145
	28C	2.283	2.361	2.776	2.626	0.194	0.123	0.130	0.229	0.121	0.042	0.371	0.222
MPSL	6C	2.395	2.267	2.846	2.667	0.221	0.161	0.076	0.203	0.279	0.407	0.171	0.007
	28C	2.663	2.667	2.913	2.844	0.191	0.172	0.152	0.204	0.081	0.085	0.332	0.263
IPSL	6C	2.443	2.490	2.742	2.547	0.199	0.186	0.104	0.181	0.152	0.106	0.145	0.048
	28C	2.503	2.561	2.772	2.601	0.209	0.229	0.154	0.264	0.088	0.030	0.180	0.089
IDSL	6C	2.433	2.440	2.691	2.486	0.110	0.113	0.077	0.151	0.127	0.120	0.130	0.075
	28C	2.437	2.490	2.727	2.622	0.163	0.158	0.104	0.169	0.208	0.191	0.045	0.059

Table S10 Mean, width and Δ mean of the modelled distances for both MMM (ambient and cryogenic temperature) and MtsslWizard (Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each labelled with the four nitroxides (MTSL, MPSL, IPSL and IDSL), based on the ESMFold structure

1.10 Correlation plots and mean, widths values of the modelled distributions

Correlation plots (**Fig. S13** and **Fig. S14**) were obtained by extracting the mean and width of every distance distribution from the experimental and the *in silico* labelling data (**Table S5**). To get a numerical quantification of the curve discrepancies, the differences between the experimental and *in silico* mean values were also extracted as absolute values (Δ mean) (**Tables S8-S11**). The correlation plots were built considering the mean of the experimental distances on the x-axis and the mean of the *in silico* labelling on the y-axis. The widths of the experimental and simulated distributions were employed as "error bars" for each data point.

AlphaFold2		Mean				Width				Δ Mean			
		MMM		MtsslWizard		MMM		MtsslWizard		MMM		MtsslWizard	
		T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose
MTSL	6C	2.349	2.425	2.720	2.517	0.089	0.123	0.074	0.184	0.052	0.023	0.318	0.115
	28C	2.108	2.217	2.746	2.606	0.196	0.220	0.152	0.235	0.295	0.186	0.341	0.202
MPSL	6C	2.364	2.213	2.852	2.643	0.223	0.123	0.070	0.195	0.310	0.460	0.177	0.031
	28C	2.525	2.499	2.930	2.795	0.173	0.173	0.132	0.229	0.056	0.082	0.348	0.213
IPSL	6C	2.365	2.473	2.718	2.528	0.223	0.197	0.115	0.199	0.232	0.123	0.121	0.068
	28C	2.525	2.408	2.768	2.659	0.173	0.217	0.140	0.216	0.066	0.183	0.175	0.067
IDSL	6C	2.436	2.437	2.645	2.478	0.115	0.115	0.079	0.159	0.124	0.123	0.084	0.082
	28C	2.331	2.339	2.714	2.591	0.148	0.148	0.105	0.176	0.350	0.342	0.028	0.090

Table S8 Mean, width and Δ mean of the modelled distances for both MMM (ambient and cryogenic temperature) and MtsslWizard (Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each labelled with the four nitroxides (MTSL, MPSL, IPSL and IDSL), based on the AlphaFold2 structure

X-ray		Mean				Width				Δ Mean			
		MMM		MtsslWizard		MMM		MtsslWizard		MMM		MtsslWizard	
		T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose	T_{cryo}	T_{amb}	Tight	Loose
MTSL	6C	2.304	2.391	2.703	2.472	0.127	0.148	0.084	0.191	0.097	0.010	0.300	0.070
	28C	2.193	2.151	2.689	2.552	0.163	0.263	0.157	0.257	0.210	0.252	0.285	0.148
MPSL	6C	2.307	2.319	2.822	2.634	0.219	0.176	0.076	0.188	0.367	0.355	0.147	0.040
	28C	2.519	2.272	2.860	2.721	0.171	0.125	0.186	0.269	0.062	0.307	0.278	0.139
IPSL	6C	2.394	2.477	2.691	2.508	0.228	0.183	0.108	0.195	0.201	0.118	0.094	0.088
	28C	2.351	2.394	2.673	2.522	0.217	0.233	0.173	0.271	0.240	0.198	0.080	0.069
IDSL	6C	2.413	2.428	2.615	2.454	0.113	0.109	0.063	0.170	0.148	0.133	0.054	0.106
	28C	2.332	2.346	2.620	2.518	0.166	0.158	0.115	0.188	0.349	0.335	0.061	0.164

Table S11 Mean, width and Δ mean of the modelled distances for both MMM (ambient and cryogenic temperature) and MtsslWizard (Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each labelled with the four nitroxides (MTSL, MPSL, IPSL and IDSL), based on the X-ray crystallographic structure (PDB: 4wh4)

A first set of correlation plots investigates the dependence on the labelling approach (MtsslWizard Tight or Loose and MMM ambient or cryogenic temperature) of the four different nitroxide labels for the same structure prediction method.

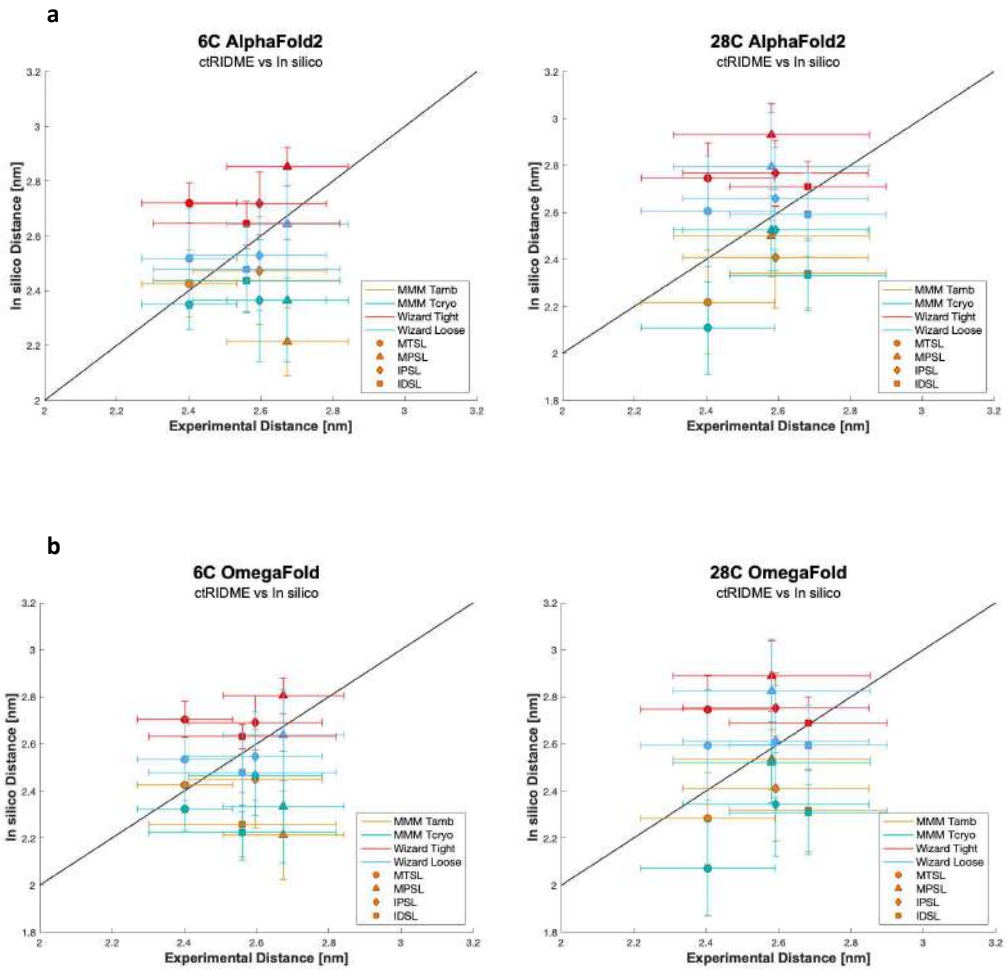


Fig. S13 Correlation plots, each point was plotted considering the mean value of the experimental distance distribution of the ctRIDME for the x-axis and their respective *in silico* mean values on the y axis. The error bars of each data point are derived from the experimental distance distribution width on the x axis and the respective *in silico* width on the y axis. The black line scores for the experimental trend. The different colors represent the different labelling approaches at different conditions: MMM at ambient temperature (orange), MMM at cryogenic temperature (green), MtsslWizard with Tight settings (red), MtsslWizard with Loose settings (blue). The different shapes correspond to the different nitroxide labels: circle MTSL, triangle MPSL, diamond IPSL and square IDSL. **a)** AlphaFold2, **b)** OmegaFold, **c)** ESMFold, **d)** X-ray (PDB: 4wh4)

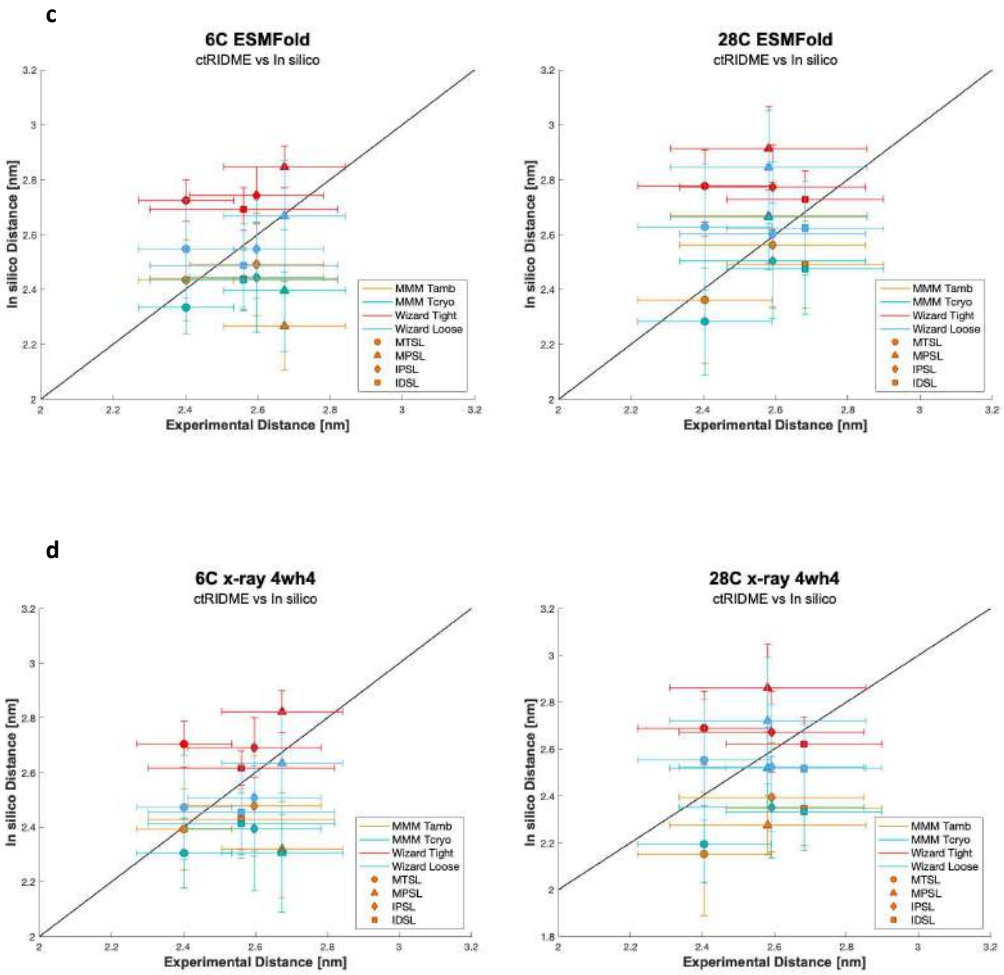


Fig. S13, continued Correlation plots, each point was plotted considering the mean value of the experimental distance distribution of the ctRIDME for the x-axis and their respective *in silico* mean values on the y axis. The error bars of each data point are derived from the experimental distance distribution width on the x axis and the respective *in silico* width on the y axis. The black line scores for the experimental trend. The different colors represent the different labelling approaches at different conditions: MMM at ambient temperature (orange), MMM at cryogenic temperature (green), MtsslWizard with Tight settings (red), MtsslWizard with Loose settings (blue). The different shapes correspond to the different nitroxide labels: circle MTSL, triangle MPSL, diamond IPSL and square IDSL. **a)** AlphaFold2, **b)** OmegaFold, **c)** ESMFold, **d)** X-ray (PDB: 4wh4)

A second set of correlation plots was introduced to better understand the role of the structure prediction methods with respect to the different labelling approaches for the same nitroxide label on the same GB1 construct.

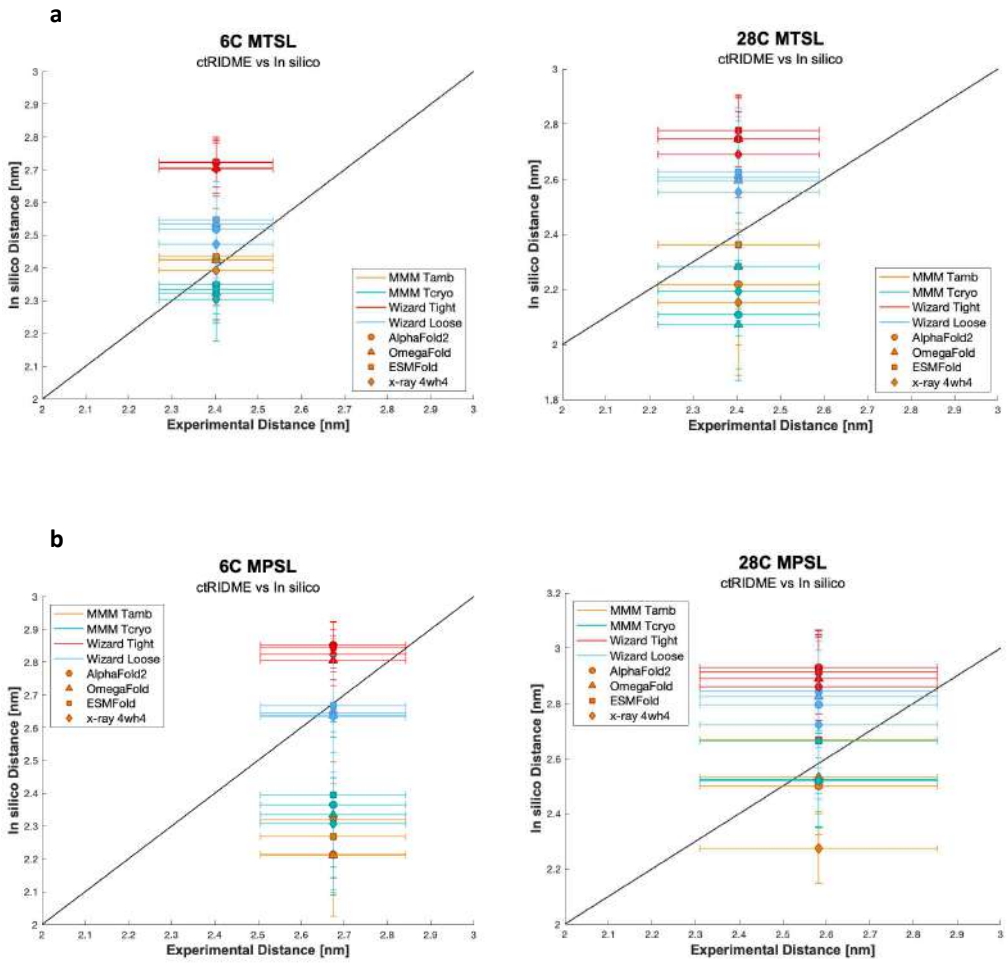


Fig. S14 Correlation plots, each point was plotted considering the mean value of the experimental distance distribution of the ctRIDME for the x-axis and their respective *in silico* mean values on the y axis. The error bars of each data point derive from the experimental distance distribution width on the x axis and the respective *in silico* width on the y axis. The black line scores for the experimental trend. The different colors represent the different labelling approaches at different conditions: MMM at ambient temperature (orange), MMM at cryogenic temperature (green), MtsslWizard with Tight settings (red), MtsslWizard with Loose settings (blue). The different shapes correspond to the different structure predictor tools: circle AlphaFold2, triangle OmegaFold, diamond X-ray (PDB: 4wh4) and square ESM-Fold. **a)** 6C and 28C MTSL, **b)** 6C and 28C MPSL, **c)** 6C and 28C IPSL, **d)** 6C and 28C IDSL.

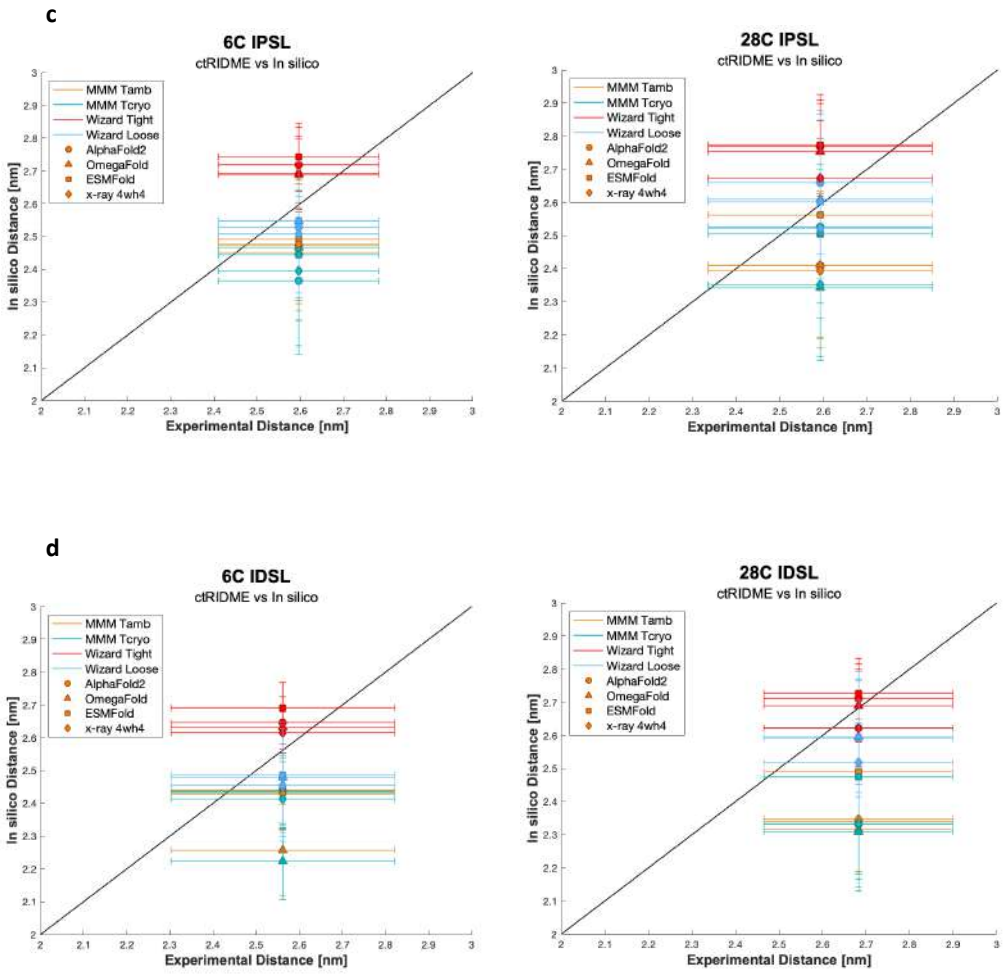


Fig. S14, continued Correlation plots, each point was plotted considering the mean value of the experimental distance distribution of the ctRIDME for the x-axis and their respective *in silico* mean values on the y axis. The error bars of each data point derive from the experimental distance distribution width on the x axis and the respective *in silico* width on the y axis. The black line scores for the experimental trend. The different colors represent the different labelling approaches at different conditions: MMM at ambient temperature (orange), MMM at cryogenic temperature (green), MtsslWizard with Tight settings (red), MtsslWizard with Loose settings (blue). The different shapes correspond to the different structure predictor tools: circle AlphaFold2, triangle OmegaFold, diamond X-ray (PDB: 4wh4) and square ESM-Fold. **a)** 6C and 28C MTSL, **b)** 6C and 28C MPSL, **c)** 6C and 28C IPSL, **d)** 6C and 28C IDSL.

1.11 Root mean square deviation (RMSD) values

To compare the overall shapes of the experimental and *in silico* distance distributions, the rmsd values were calculated with an in-house Matlab software (**Tables S12-S15**).

AlphaFold2

RMSD		MTSL		MPSL		IPSL		IDSL	
		6C	28C	6C	28C	6C	28C	6C	28C
MMM	T_{cryo}	0.0070	0.0159	0.0132	0.0072	0.0118	0.0082	0.0094	0.0193
	T_{amb}	0.0083	0.0149	0.0202	0.0075	0.0068	0.0081	0.0094	0.0191
Wizard	Tight	0.0247	0.0197	0.0182	0.0082	0.0121	0.0082	0.0190	0.0096
	Loose	0.0153	0.0170	0.0082	0.0057	0.0057	0.0057	0.0135	0.0065

Table S12 Root mean square deviation (rmsd) values of the modelled distances for both MMM and MtsslWizard at both different conditions (ambient and cryogenic temperature and Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each respectively with the four nitroxides (MTSL, MPSL, IPSL and IDSL), for the AlphaFold2 structure

OmegaFold

RMSD		MTSL		MPSL		IPSL		IDSL	
		6C	28C	6C	28C	6C	28C	6C	28C
MMM	T_{cryo}	0.0098	0.0169	0.0133	0.0074	0.0071	0.0088	0.0226	0.0190
	T_{amb}	0.0076	0.0147	0.0175	0.0121	0.0067	0.0079	0.0201	0.0188
Wizard	Tight	0.0245	0.0197	0.0135	0.0117	0.0099	0.001	0.0228	0.0070
	Loose	0.0166	0.0158	0.0075	0.0101	0.0057	0.0046	0.0130	0.0065

Table S13 Root mean square deviation (rmsd) values of the modelled distances for both MMM and MtsslWizard at both different conditions (ambient and cryogenic temperature and Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each respectively with the four nitroxides (MTSL, MPSL, IPSL and IDSL), for the OmegaFold structure

ESMFold

RMSD		MTSL		MPSL		IPSL		IDSL	
		6C	28C	6C	28C	6C	28C	6C	28C
MMM	T_{cryo}	0.0085	0.0137	0.0123	0.0058	0.0069	0.0066	0.0104	0.0118
	T_{amb}	0.0114	0.0139	0.0176	0.0055	0.0067	0.0066	0.0098	0.0108
Wizard	Tight	0.0246	0.0205	0.017	0.0120	0.0135	0.0078	0.0224	0.0102
	Loose	0.0165	0.0166	0.0084	0.0101	0.0057	0.0051	0.0120	0.0070

Table S14 Root mean square deviation (rmsd) values of the modelled distances for both MMM and MtsslWizard at both different conditions (ambient and cryogenic temperature and Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each respectively with the four nitroxides (MTSL, MPSL, IPSL and IDSL), for the ESMfold structure

X-ray 4wh4

RMSD		MTSL		MPSL		IPSL		IDSL	
		6C	28C	6C	28C	6C	28C	6C	28C
MMM	T_{cryo}	0.0099	0.0137	0.0151	0.0086	0.0077	0.009	0.0115	0.0184
	T_{amb}	0.0106	0.0139	0.0169	0.0133	0.0075	0.0081	0.0100	0.0185
Wizard	Tight	0.0239	0.0189	0.0036	0.0128	0.0103	0.0059	0.0201	0.0034
	Loose	0.0146	0.0165	0.0104	0.081	0.0051	0.0055	0.0129	0.0074

Table S15 Root mean square deviation (rmsd) values of the modelled distances for both MMM and MtsslWizard at both different conditions (ambient and cryogenic temperature and Tight and Loose settings), for both GB1 constructs I6C/K28H/Q32H and I6H/N8H/K28C, each respectively with the four nitroxides (MTSL, MPSL, IPSL and IDSL), for the X-ray crystallographic structure (PDB: 4wh4)

To better estimate which labelling approach globally predicts more consistently the experimental behaviour, we compared at the same time all the mean values of the distance distributions of both constructs with the four distinct nitroxide spin labels for a single *in silico* labelling approach, with the same experimental data, extracting the global rmsd values (**Table S16**). This procedure was repeated for the two labelling approaches and their respective conditions (ambient and cryogenic temperatures for MMM and Tight and Loose settings for MtsslWizard). All three different structure prediction methods (AlphaFold2, OmegaFold and ESMFold) and the X-ray crystallographic structure were analysed. The same procedure was employed to analyse the distribution widths (**Table S16**).

RMSD Global		AlphaFold2		OmegaFold		ESMFold		X-ray	
		Mean	Width	Mean	Width	Mean	Width	Mean	Width
MMM	T_{cryo}	0.219	0.077	0.267	0.071	0.156	0.069	0.233	0.071
	T_{amb}	0.233	0.071	0.252	0.071	0.173	0.068	0.242	0.082
Wizard	Tight	0.230	0.110	0.211	0.112	0.238	0.107	0.191	0.100
	Loose	0.125	0.051	0.127	0.048	0.137	0.054	0.111	0.047

Table S16 Global RMSD values, extracted by comparison between all the mean values (for both GB1 constructs and the four nitroxide labels) for a single labelling approach with the experimental values. Repeated for the two labelling approaches MMM and MtsslWizard with their respective conditions, cryogenic and ambient temperature for the former, and Tight and Loose settings for the latter. All predicted structures (AlphaFold2, OmegaFold, ESMFold) and the X-ray crystallographic structure have been considered. The same procedure was employed to analyse the distribution widths

II References

[1] K. Ackermann, J.L. Wort, B.E. Bode, Pulse dipolar EPR for determining nanomolar binding affinities, *Chem Commun,* **2022**; 58, 8790-8793, https://doi.org/10.1039/d2cc02360a.

[2] S. Ghosh, S. Saxena, G. Jeschke, Rotamer Modelling of Cu(II) Spin Labels Based on the Double-Histidine Motif, *Applied Magnetic Resonance,* **2018**; 49, 1281-1298, https://doi.org/10.1007/s00723-018-1052-8.

5.3. A systematic study on the effect of protonation and deuteration on T_m in a cellular context

In-cell EPR is still a relatively young field where several improvements can still be made, from the labels stability to the protein delivery protocols with the aim of increasing the sensitivity of the technique. In this work, for the first time we investigated the effect of deuterated cells, both prokaryotic (*Escherichia coli*) and eukaryotic (*Pichia pastoris*) on the protein relaxation time T_m. Deuterated cells can be extremely helpful in increasing the phase memory time both for retrieving longer distances between paramagnetic labels and for overcoming the poor signal to noise ratio derived from the generally low concentration of the in-cell samples. Here we expressed both fully deuterated and protonated doubly cysteine mutated human ubiquitin (S20C/G35C), which were delivered into cells through the thermal delivery protocol, after the labelling reaction with the malemide proxyl spin label (MPSL). At the same time, both protonated and deuterated *Escherichia Coli* and *Pichia Pastoris* - the latter obtained through a novel protocol for the preparation of deuterated *Pichia pastoris* cells - were obtained and used for the thermal delivery. The samples of protonated protein in protonated cells, deuterated protein in protonated cells, protonated protein in deuterated cells and deuterated protein in deuterated cells were analyzed through EPR spectroscopy. Their corresponding T_m were evaluated to check which experimental protein/cell combination could be the most suitable considering T_m gaining but also the global experimental costs. Before the in cell experiments the T_m characterization of the doubly labelled protein were conducted in buffer solution. Indeed, four sample of the protein in different environment were properad as follow: protonated protein in protonated solvent, protonated protein in deuterated solvent, deuterated protein in protonated solvent and deuterated protein in deuterated sample. The T_m values were estimated by fitting the experimental data with a stretched exponential equation and the results clearly confirm how the deuteration highly affects and increases the overall T_m of the labelled protein (where in the fully deuterated case the T_m was estimated to be more than doubled compared to the fully protonated case). The in-cell samples were prepared using the previously introduced thermal stimulation delivery (for both deuterated and protonated cells). For the *E. coli*, we proved how the fully deuterated cells can extend the relative T_m of the protonated

protein from 2.0 ± 0.3 to 8.7 ± 1.1 μs, while in the case of the deuterated protein the value is increased from 1.4 ± 0.2 to 12.0 ± 0.4 μs in the deuterated cells. Even though the use of fully deuterated sample, in terms of protein and cells gives the best increasing in T_m, it also represents the most expensive choice. For this reason, in the *E. coli* case, the use of fully protonated protein delivered in fully deuterated cells gives a less expensive way to get a good enhancement of the T_2 in the in-cell environment. Regarding the *P. pastoris* using fully deuterated cells still gives an enhancement of the relative T_m even if the overall effects is less impressive. Indeed, we showed how the T_m is increased from 3.0 ± 0.3 to 4.6 ± 0.1 μs for the protonated ubiquitin delivered in protonated and deuterated cells. On the same line, the T_m values were shown to be increased from 4.9 ± 0.2 to 5.8 ± 0.5 μs for the deuterated protein delivered inside protonated and deuterated *P. pastoris* cells. Based on these latter results we can speculate how the more complex and compartmentalized intracellular environment of the *P. pastoris* can lead to an additional effect on the overall measured T_m values which can decrees the effects of the using deuterated systems for increasing its value. Nevertheless, the effect of deuteration is still evident and the even in this case the best choice in terms of balancing the cost pf the experiments and gaining a substantial gain in T_m is still working with protonated protein in deuterated cells.

COMMUNICATION

A systematic study on the effect of protonation and deuteration on electron spin T_m in a cellular context

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Francesco Torricella[a,b], Valentina Vitali[a] and Lucia Banci[a,c,d*].

Over the past decade, the DEER experiment in pulsed EPR has gained attention for its unique ability to provide precise distance distribution information in cellular and buffer environments. This method is closely tied to the sample's transverse relaxation time T_m (or T_2), limiting access to long-distance distributions. Our study demonstrates how using either fully protonated or deuterated cellular environments can significantly influence the transverse relaxation time. Notably, combining deuterated cells with protonated spin-labeled proteins greatly enhances T_m relaxation times within our spectroscopic setup. This approach improves the accuracy and versatility of DEER experiments for investigating molecular distances and structures in various contexts.

The pulsed electron-electron double resonance (PELDOR or DEER) technique can be utilized as a unique pulse-EPR method for accurately retrieving the distance distribution between two unpaired electrons[1]. Typically, these electrons are spatially located on specific labels (e.g., nitroxide[2], metal label[2, 3], or trityl radicals[4]) directly connected to specific residues like Cys[5] or Tyr[6, 7]. In the last decades, the use of the DEER sequences has been extended from the regular buffer sample, where the doubly labelled protein is in a more or less complex buffer solutions, to a cellular environment to the so-called in-cell experiments[8-16]. DEER experiment has proven to be a background-free method for investigating protein structure, dynamics, and interactions within the intracellular environment of different host cells[10, 17-20]. Similarly, it has been demonstrated how specifically designed nitroxide labels can overcome the bio-reduction of the regular commercially available nitroxides inside the intracellular environment employing shielding groups nearby the nitroxide motif[9, 21-23]. Indeed, we previously showed how the maleimide Proxyl (ma-Px) bio-reduction lifetime can be exploited for in-cell investigation with the use of an appropriate delivery protocol with the previous mentioned ma-Px[9, 17, 18]. In both buffer (referred to as *in vitro* in this article) and in-cell experiments with nitroxide-based radicals, the typical DEER experiment is conducted at a cryogenic temperature of 50 K using a frozen glassy sample containing different percentages of glycerol. These conditions are chosen so to maximize the transverse relaxation time T_2/T_m and the longitudinal relaxation rate T_1, opening the possibility of using long time delays in the spin echo sequence coupled with high repetition rates for acquisition. Using these long spin echo delays gives the possibility to retrieve relatively long distances between the two selected paramagnetic centers[24]. Considering an interelectron spin distance of 100 Å, the expected nutation time that one would need to wait for gathering this distance distribution needs to be at least equal to 20 μs[24].

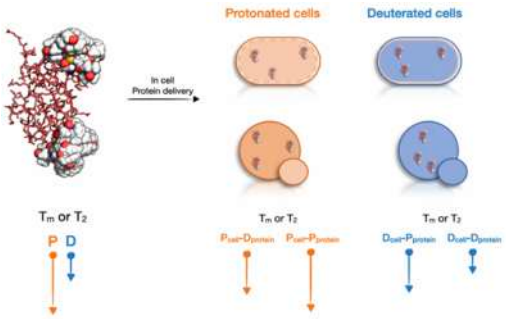


Figure 1: (A) Graphical representation of the presented study. On the left ubiquitin structure with the computed rotamers for the two ma-Px molecules (MMM model)[25]. On the right a representation of protonated (orange) and deuterated (blue) cells. First row represents the prokaryotic *E.coli* cells, while the second stands for the eukaryotic Pichia Pastoris cells. The T_2 values for the tagged ubiquitin have been measured after its delivery to either fully protonated or deuterated *Escherichia coli* and *Pichia Pastoris* cells. Length of the below arrows graphically represents the overall decay velocity of T_2 in the time dimension. Long arrows stand for shorter T_2. As shown by the graphical representation the electron T_2 is strictly connected to the relative environment.

In a regular protonated sample, the sequence delays for catching this kind of nutation time cannot be accessed because of the relative

[a] Magnetic Resonance Center, Università degli Studi di Firenze, 50019 Sesto Fiorentino, Italy.
[b] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health, Bethesda, Maryland 20892-0520, United States.
[c] Consorzio Interuniversitario Risonanza Magnetiche di Metallo Proteine, 50019 Sesto Fiorentino, Italy.
[d] Dipartimento di Chimica, Università degli Studi di Firenze, 50019 Sesto Fiorentino, Italy.

short T_2. In fact, it is not uncommon that the electron spin T_2 of the sample is faster than the needed nutation time, thus disrupting the possibility of getting the desired distance distribution. These nutation times and the related DEER delays are unobtainable in a regular fully protonated biological sample. To overcome this problem several papers have suggested the potentialities of the use of deuteration[24, 26-28]. Nowadays, solvent deuteration represents the standard practice to increase the relaxation rates from the effects that concern the solvent[29, 30]. However, a way stronger approach for increasing the overall electron relaxation times is based on the use of fully deuterated proteins[28]. This condition can extend dramatically the electron spin T_2 relaxation time, so giving the possibility of measuring longer distance distributions between pair of spin labels[24]. Furthermore, another advantage of slower relaxation rates is the relative higher sensitivity, that obviously improves the data acquisition times and the ability to gather data with lower concentrated samples. In this regard, one of the main limitations of the in-cell experiments is represented by the overall concentration reached inside the intracellular environment after the protein delivery. The protein concentration could be significantly increased if the in-cell sample preparation is changed from delivery to overexpression-based protocols[31, 32].In this work we report on a systematic investigation of the effects of the use of fully protonated and deuterated samples on electronic T_2. Particularly, we employed protonated and deuterated doubly labelled human Ubiquitin (Ub), which was delivered in both prokaryotic (*Escherichia Coli*) and eukaryotic (*Pichia Pastoris*) cells grown in either fully protonated or fully deuterated media.

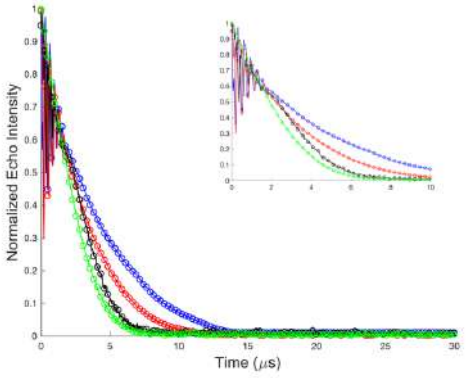


Figure 2: Echo decay curves of the four-samples of the doubly labeled ubiquitin in buffer. In particular: green curve is from the protonated protein in protonated buffer, black curve from the protonated protein in deuterated buffer, red curve from the deuterated protein in protonated buffer and the blue curve from the deuterated protein in deuterated buffer.

The overall idea behind this work is the investigation of the effects on electronic spin T_2 induced by the selective use of different combinations of protonation and deuteration of the target protein and of the cells for in-cell EPR experiments with the use of thermal delivery[18]. To do so, we selected the double cysteine (S20C/G35C) mutated human ubiquitin, doubly labelled with ma-Px and two different cell hosts. First, we analyzed the electron spin T_2 properties of the doubly labelled protein in buffer solutions. Four samples of the protein were prepared as follow: 1) protonated protein in protonated buffer, 2) protonated protein in deuterated buffer, 3) deuterated protein in protonated buffer, and 4) deuterated protein in deuterated buffer. The electron spin T_2 values were measured with the standard two pulse echo decay sequence following the (π/2–τ–π– τ –echo) pulse sequence. Echo decays curves, (**Figure 2),** show, as already reported in literature[21], that both protein and solvent deuteration produces an increase of the electron spin T_2 of the labelled protein. The T_2 values were calculated by fitting the experimental echo decay data with a stretched exponential equation (Equation 1).

$$y = a\, e^{-\left(\frac{x}{b}\right)^c} + d \quad (1)$$

From the data fitting the electron spin T_2 for the fully protonated sample is found to be 2.6 ± 0.2 μs, while, for the protonated protein in deuterated buffer is 3.4 ± 0.1 μs which increases to 4.0 ± 0.2 μs in the case of the deuterated protein in protonated buffer. The fully deuterated sample has an electron spin T_2 of 5.5 ± 0.3 μs, i.e. more than doubled of that of the fully protonated sample, consistent with the expected effects of protein and buffer deuteration effects on the relaxation rates. Notably, in our system the overall effect, even if significant, is not very high due to the instrument set-up, i.e. relatively low nominal pulse power, to the intrinsic protein characteristic, to the labeling position and mainly to the glycerol content in the sample. In fact, as glycerol is needed for the in-cell samples for ensuring cell stability upon sample freezing, 10% v/v of glycerol has been added to all the measured samples in this work, so to have the same percentage of glycerol in the final sample and to avoid any possible effect due to different concentrations of glycerol, which at huge percentage affects the T_2[33]. We then measured electron spin T_2 values of the labelled ubiquitin in-cell to see the effect of the fully intracellular deuterated environment to the electron spin T_2 of the delivered protein. *Escherichia coli* (E. coli) cells and *Pichia Pastoris* (P. pastoris) cells have been selected. As extensively shown for NMR applications, E. coli cells can be used to produce fully deuterated proteins and therefore, we were able to grow E. coli in a fully deuterated medium[34]. On the other hand, for the eukaryotic cells P. Pastoris we developed a novel protocol for producing deuterated cells from a specific fully deuterated growing medium (Supporting Information). All in-cell samples, either with protonated or deuterated cells, were prepared using the thermal stimulation delivery[18] with a protein external concentration of 500 μM. All the preparative steps along with the cell preparation protocols are reported in the Material and Methods section and Supplementary Information. The echo decay curve acquired for the in-cell *E.coli* samples are shown in **Figure 2A**. The additional acquired X-band cw-EPR spectra acquired at RT and the relative ESE spectra acquired at Q-band are summarized in **Figure S1**. From the data fitting (**Figure S3**), the electron spin T_2 values resulted to be 2.0 ± 0.3

μs for the protein inside the protonated cells and 8.7 ± 1.1 μs for that inside the fully deuterated cells. Additionally, electron spin T_2 was found to be 1.4 ± 0.2 μs for the deuterated protein delivered inside protonated cells and 12.0 ± 0.4 μs for the deuterated protein inside deuterated cells. As introduced above, the measured electron spin T_2 is dramatically affected by the quantity of glycerol in the used buffer, as previously reported in the literature by Clore et al.[33] . To extend the investigation towards more challenging systems such as eukaryotic cells, as introduced Pichia Pastoris was selected. We developed a novel strategy for growing yeast cells in a fully deuterated media which allowed the preparation of fully deuterated P. Pastoris cells for the thermal stimulation delivery. Indeed, even for this cell host, we prepared four in-cell samples using the same external protein concentration and the same delivery set-up as used for *E. coli* (Supporting Information). All the acquired decay curves are shown in **Figure 3A** and **3B**.

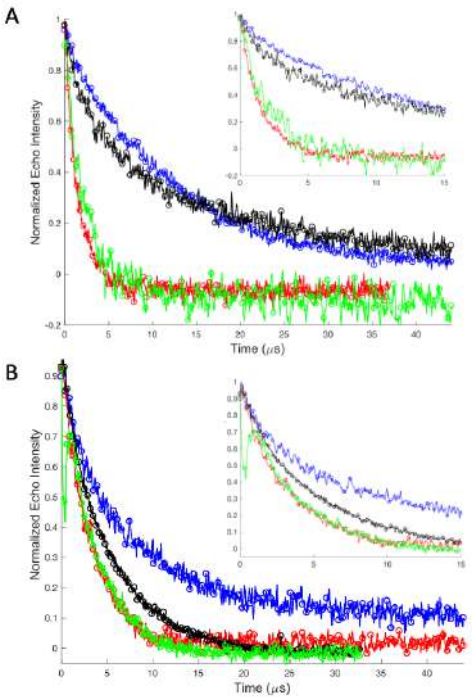


Figure 3: (A) Echo decay curves for all the four-sample acquired in E.coli cells for the doubly labeled ubiquitin. In particular: green curve was acquired from for the sample with protonated protein in protonated cells, black curve was acquired from the sample with protonated protein in deuterated cells, red curve was acquired from the deuterated protein in protonated cells and the blue curve was acquired from the deuterated protein in deuterated cells. (B) Echo decay curves for all the four-sample acquired in P.pastoris cells for the doubly labeled ubiquitin. In particular: green curve was acquired from for the sample with protonated protein in protonated cells, black curve was acquired from the sample with protonated protein in deuterated cells, red curve was acquired from the deuterated protein in protonated cells and the blue curve was acquired from the deuterated protein in deuterated cells.

From the data analysis the electron spin T_2 values showed the same trend as for the *E. coli* samples. In fact, we derived the following estimated T_2 value, 3.0 ± 0.3 μs for protonated ubiquitin inside the protonated cells, 4.6 ± 0.1 μs for the deuterated protein into protonated cells, 4.9 ± 0.2 μs for the protonated protein inside deuterated cells and, 5.8 ± 0.5 μs for the fully deuterated system (**Figure S4**). The herein presented data show the impact of the protein/solvent/cell deuteration on the electron spin T_2 relaxation time which can have high relevance for extending the lifetime of the electron spin transverse magnetization thus allowing the measurement of longer spin-spin distances. The T_2 relaxation parameter is in fact essential for PDS-EPR measurements like DEER spectroscopy. Indeed, longer delays in the DEER sequence can be employed with adequate T_2 values, giving the chance of tracking longer distances. Notably as expected, replacing protein protons with deuterons results in an increase of the T_2 value. As reported in **Table 1**, our data show how the transverse relaxation rates are strongly affected by the presence or absence of deuterons. For the in vitro case, the deuterated protein in the deuterated buffer showed the longest T_2 value compared to the other conditions. It is important to remark that this value can be extended even further using higher percentage of glycerol in the in vitro buffer samples, which however cannot be applied to the in-cell samples, based on the absence of the intime contact between the resuspension buffer and the labelled protein. For the in-cell samples, for *E. coli*, we showed how the fully deuterated cells can make the electron spin T_2 of the protonated protein longer, from 2.8 μs to 8.0 μs, while in the case of the deuterated protein the value is increased from 1.4 μs in protonated cells to 12.0 μs in the deuterated cells. Even though the use of fully deuterated samples, i.e. both deuterated protein and cells, gives the largest increase in the electron spin T_2, it also represents the most expensive choice. On this respect, for E. coli cells, the use of protonated protein in fully deuterated cells gives a less expensive way to get a sizable increase of the T_2. In *P. pastoris*, the use of fully deuterated cells still gives an increase of the T_2 value even if the overall effect is lower. Indeed, the T_2 value is increased from 3.0 to 4.6 μs for protonated ubiquitin delivered to protonated vs deuterated cells. On the same line, the electron spin T_2 relaxations times increase from 4.9 to 5.8 μs for the deuterated protein delivered inside protonated and deuterated P. pastoris cells. Based on these latter results we can speculate that the more complex and compartmentalized intracellular environment of P. pastoris could have further, opposite effects on the overall measured T_2 values, which could reduce the effects of using deuterated systems.

Sample	T_2 (μs)	Sample	T_2 (μs)	Sample	T_2 (μs)	Sample	T_2 (μs)	Sample	T_2 (μs)
Ub_{pp}	H 2.6 ± 0.02 D 5.4 ± 0.1	$E.coli_{pp}$	2.0 ± 0.3	$E.coli_{pd}$	8.7 ± 1.1	$P.pastoris_{pp}$	3.0 ± 0.3	$P.pastoris_{pd}$	4.6 ± 0.1
Ub_{pd}	H 4.0 ± 0.2 D 5.5 ± 0.3	$E.coli_{pp}$	1.4 ± 0.2	$E.coli_{pd}$	12.0 ± 0.4	$P.pastoris_{pp}$	4.9 ± 0.2	$P.pastoris_{pd}$	5.8 ± 0.5

Table 1: Estimated values for T2

Conclusions

In conclusion, this work showed for the first time how the selective use of protein protonation and deuteration combined with the use of protonated and deuterated cells can be used to increase the overall value of the electron spin T_2, a relevant feature for measuring longer spin-spin distances in the intracellular environment. As showed from the above reported values the best approach in term of T_2 increase and overall relative cost for the experiments is the use of protonated proteins and deuterated cellular hosts, as clearly showed by the experiments in E.coli. In addition, the preparation of deuterated proteins can be sometimes cumbersome based on low yield for the expression or even the absence of protein expression in the deuterate medium. At the same time, the preparation of fully deuterated cells represents a way easier procedure with a higher success rate. The measurement of long spin-spin distances in-cell would allow the intracellular investigation of large complexes via EPR and pulsed EPR methods, which can be groundbreaking in structural biology due to its unique contribution and absence of any molecular weight limitations.

Author Contributions

F.T theorized the experiments. F.T L.B designed the experiments. F.T, V.V and L.B acquired the data and analysed them. All the authors contribute to write and review the paper.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

This work was supported by Instruct-ERIC, a Landmark ESFRI project, and specifically by the CERM/CIRMMP Italian Instruct Center.

Notes and references

1. O. Schiemann and T. F. Prisner, *Quarterly reviews of biophysics*, 2007, **40**, 1-53.
2. V. Vitali, K. Ackermann, G. Hagelueken and B. E. Bode, *In Dataset*, 2023, **University of St Andrews Research Portal**.
3. A. Giannoulis, Y. Ben-Ishay and D. Goldfarb, *Methods Enzymol*, 2021, **651**, 235-290.
4. N. Fleck, C. A. Heubach, T. Hett, F. R. Haege, P. P. Bawol, H. Baltruschat and O. Schiemann, *Angew Chem Int Ed Engl*, 2020, **59**, 9767-9772.
5. F. Torricella, A. Pierro, E. Mileo, V. Belle and A. Bonucci, *Biochim Biophys Acta Proteins Proteom*, 2021, **1869**, 140653.
6. E. Mileo, E. Etienne, M. Martinho, R. Lebrun, V. Roubaud, P. Tordo, B. Gontero, B. Guigliarelli, S. R. A. Marque and V. Belle, *Bioconjugate Chemistry*, 2013, **24**, 1110-1117.
7. V. Vitali, F. Torricella, L. Massai, L. Messori and L. Banci, *Scientific Reports*, 2023, **13**, 22017.
8. A. Bonucci, O. Ouari, B. Guigliarelli, V. Belle and E. Mileo, *Chembiochem*, 2020, **21**, 451-460.
9. A. Pierro, K. C. Tamburrini, H. Leguenno, G. Gerbaud, E. Etienne, B. Guigliarelli, V. Belle, B. Zambelli and E. Mileo, *iScience*, 2023, **26**, 107855.
10. A. Pierro, A. Bonucci, D. Normanno, M. Ansaldi, E. Pilet, O. Ouari, B. Guigliarelli, E. Etienne, G. Gerbaud, A. Magalon, V. Belle and E. Mileo, *Chemistry – A European Journal*, 2022, **28**, e202202249.
11. Y. Yang, B. B. Pan, X. Tan, F. Yang, Y. Liu, X. C. Su and D. Goldfarb, *J Phys Chem Lett*, 2020, **11**, 1141-1147.
12. Y. Yang, S. N. Chen, F. Yang, X. Y. Li, A. Feintuch, X. C. Su and D. Goldfarb, *Proc Natl Acad Sci U S A*, 2020, **117**, 20566-20575.
13. A. Dalaloyan, A. Martorana, Y. Barak, D. Gataulin, E. Reuveny, A. Howe, M. Elbaum, S. Albeck, T. Unger, V. Frydman, E. H. Abdelkader, G. Otting and D. Goldfarb, *Chemphyschem*, 2019, **20**, 1860-1868.
14. Y. Yang, F. Yang, X. Y. Li, X. C. Su and D. Goldfarb, *J Phys Chem B*, 2019, **123**, 1050-1059.
15. Y. Shenberger, L. Gevorkyan-Airapetov, M. Hirsch, L. Hofmann and S. Ruthstein, *Chem Commun (Camb)*, 2023, **59**, 10524-10527.
16. D. Goldfarb, *Current Opinion in Structural Biology*, 2022, **75**, 102398.
17. F. Torricella, L. Barbieri, V. Bazzurro, A. Diaspro and L. Banci, *Sci Rep*, 2022, **12**, 17190.
18. F. Torricella, A. Bonucci, P. Polykretis, F. Cencetti and L. Banci, *Biochem Biophys Res Commun*, 2021, **570**, 82-88.
19. A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime and D. Goldfarb, *J Am Chem Soc*, 2014, **136**, 13458-13465.
20. Y. Yang, F. Yang, Y. J. Gong, T. Bahrenberg, A. Feintuch, X. C. Su and D. Goldfarb, *J Phys Chem Lett*, 2018, **9**, 6119-6123.
21. S. Bleicken, T. E. Assafa, H. Zhang, C. Elsner, I. Ritsch, M. Pink, S. Rajca, G. Jeschke, A. Rajca and E. Bordignon, *ChemistryOpen*, 2019, **8**, 1057-1065.
22. G. Karthikeyan, A. Bonucci, G. Casano, G. Gerbaud, S. Abel, V. Thome, L. Kodjabachian, A. Magalon, B. Guigliarelli, V. Belle, O. Ouari and E. Mileo, *Angew Chem Int Ed Engl*, 2018, **57**, 1366-1370.
23. R. Igarashi, T. Sakai, H. Hara, T. Tenno, T. Tanaka, H. Tochio and M. Shirakawa, *J Am Chem Soc*, 2010, **132**, 8228-8229.
24. T. Schmidt, M. A. Walti, J. L. Baber, E. J. Hustedt and G. M. Clore, *Angew Chem Int Ed Engl*, 2016, **55**, 15905-15909.
25. G. Jeschke, *Protein Science*, 2018, **27**, 76-85.
26. H. El Mkami and D. G. Norman, *Methods Enzymol*, 2015, **564**, 125-152.
27. H. El Mkami, R. Ward, A. Bowman, T. Owen-Hughes and D. G. Norman, *J Magn Reson*, 2014, **248**, 36-41.
28. R. Ward, A. Bowman, E. Sozudogru, H. El-Mkami, T. Owen-Hughes and D. G. Norman, *J Magn Reson*, 2010, **207**, 164-167.
29. G. Jeschke, *Chemphyschem*, 2002, **3**, 927-932.

Please do not adjust margins

30. G. Jeschke and Y. Polyhach, *Phys Chem Chem Phys*, 2007, **9**, 1895-1910.
31. P. Widder, J. Schuck, D. Summerer and M. Drescher, *Phys Chem Chem Phys*, 2020, **22**, 4875-4879.
32. A. Kugele, S. Ketter, B. Silkenath, V. Wittmann, B. Joseph and M. Drescher, *Chem Commun (Camb)*, 2021, **57**, 12980-12983.
33. T. Schmidt, J. Jeon, Y. Okuno, S. C. Chiliveri and G. M. Clore, *ChemPhysChem*, 2020, **21**, 1224-1229.
34. J. Li and R. A. Byrd, *J Biol Chem*, 2022, **298**, 102253.

Please do not adjust margins

Supporting information

A systematic study on the effect of protonation and deuteration on T_m in a cellular context

Francesco Torricella[a,b], Valentina Vitali[a] and Lucia Banci[a,c,d*].

[a] Magnetic Resonance Center, Università degli Studi di Firenze, 50019 Sesto Fiorentino, Italy.
[b] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health, Bethesda, Maryland 20892-0520, United States.
[c] Consorzio Interuniversitario Risonanza Magnetiche di Metallo Proteine, 50019 Sesto Fiorentino, Italy.
[d] Dipartimento di Chimica, Università degli Studi di Firenze, 50019 Sesto Fiorentino, Italy.

Table of Contents

1. Materials and Methods

1.1. Protein expression and purification

Protonated ubiquitin - A pET-21a vector encoding for human Ubiquitin (Ub) (containing the double mutation S20G35C) was used to transform BL21 (DE3) cell strain. E. coli cells which were grown to mid-log phase at 37° C in LB medium, and then induced with 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). After induction the cells were grown for other 5 hours at 37° C. The cells were then harvested, resuspended in lysis buffer (50 mM sodium acetate, pH 4.5) and lysed by sonication (10 seconds ON, 30 seconds OFF, at 60% of amplitude for 40 minutes). The suspension was ultracentrifuged at 40000 rpm for 40 minutes and was loaded onto a HisTrap™ SP HP column (Cytiva). Further purification was performed using size exclusion chromatography with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) exchanging the buffer with 100 mM sodium phosphate, 150 mM NaCl, 1 mM TCEP, pH 6.5.
Deuterated ubiquitin- A pET-21a vector encoding for human Ubiquitin (Ub) (containing the double mutation S20G35C) was used to transform BL21 (DE3). E. coli cells which were grown to mid-log phase at 37° C in minimal cell growth medium (6.5 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L ammonium solphate, 2 g/L D-Glucose-d12, 120 mg/L $MgSO_4$, 11 mg/L $CaCl_2$, 10 ml MEM) in deuterated water. The expression conditions and the purification steps are the same as for the protonated protein.

1.2. Cell preparation for heat shock delivery

E.Coli protonated
From a stock of 100 µL of XL1 blu competent cells a 5 mL overnight culture was grown in sterile LB medium at 37°C in a shaking incubator. The next day, the overnight culture was transferred in 500 mL of sterile LB medium. The cells were grown at 37°C until the A_{600} reached 0.35-0.4. The harvesting, washing, and storing procedures are the same than the one applied for the deuterated cells, using protonated water and not the deuterated one.

E.Coli deuterated
1 stock of XL1 blu competent cells was inoculated in 1 ml of LB in H_2O at 37°C over-day. The suspension was then transferred in 20 ml of minimal cell growth medium (6.5 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L ammonium solphate, 2 g/L D-Glucose-d12, 120 mg/L $MgSO_4$, 11 mg/L $CaCl_2$, 10 ml MEM) in deuterated water and were incubated overnight at 37°C in a shaking incubator. The next day, the incubated cells were transferred in 200 ml of the same minimal cell growth medium. The growth was carried on at 37°C to an A_{600} of 0.35-0.4. Afterwards, the culture was chilled in ice for 30 minutes and the cells were harvested by centrifugation at 3000 x *g*. The pellet was resuspended with 100mL of a deuterated ice-cold solution of 100mM $MgCl_2$, and the bacteria cells

were collected again by centrifugation at 3000 x *g*. The process was repeated with 100 mL of a pre-chilled deuterated solution of 100 mM $CaCl_2$, the *E.coli* cells were once again harvested by centrifugation at 3000 x *g*. Consecutively, bacterial cells were washed with 100 mL of a cold deuterated solution containing 85 mM of $CaCl_2$ and 15% v/v of glycerol. Finally, the cells were harvested by centrifugation at 1000 x *g*, resuspended in the aforementioned buffer, stocked in aliquots and stored at -80°C after snap freezing with liquid nitrogen.

P.Pastoris protonated
From a stock of 100 µL of X-33 *Pichia Pastoris* a 5 mL overnight culture was grown in 1 mL of YPD (10 g yeast extract, 20 g peptone, and 20 g glucose in 1L of water) at 30°C in a shaking incubator. The next day, the overnight culture was diluted to approximatively an A_{600} of 0.15–0.20 in 200 mL of YPD. The growth was carried on at 30°C to an A_{600} of 0.8-1.0, the cells were harvested by centrifugation at 500 x *g* at RT and the supernatant was poured off. The cell pellet was resuspended in PBS and stored in small aliquots of the same buffer at -80°C, after snap freezing in liquid nitrogen.

P.Pastoris deuterated
From a stock of 100 µL of X-33 *Pichia Pastoris* a 1 mL over-day culture was grown in 1 ml of YPD (10 g yeast extract, 20 g peptone, and 20 g glucose in 1L of water) at 30°C in a shaking incubator. Afterwards, the cells were transferred in 20 ml of YPD prepared with both deuterated water and D-Glucose-d12 and were incubated overnight at 30°C in a shaking incubator. The next day, the incubated cells were transferred in 200 ml of the same YPD. The growth was carried on at 30°C to an A_{600} of 0.8-1.0. Afterwards, the cells were harvested by centrifugation at 500 x *g* at RT and the supernatant was poured off. The cell pellet was resuspended in PBS deuterated and stored in small aliquots of the buffer at -80°C, after snap freezing in liquid nitrogen.

1.3. Site Direct Spin Labelling reactions and CW-EPR spectroscopy

Site directed spin labelling reactions were carried out using 3-maleimido-Proxyl label. Since the maleimide group reacts with the thiol group of cysteine, the mutations S20G35C were introduced by mutagenesis in ubiquitin. The mutants were expressed and purified following the protocols reported above. Both protonated and deuterated protein solution, were individually incubated with an excess of DTT (1:10 protein: DTT molar ratio) at RT in order to reduce the cysteine thiol group. After 1 hour, the reducing agent was removed through a PD-10 desalting column. A 10-fold molar excess of 3-maleimido-Proxyl (Sigma Aldrich), dissolved in dimethyl-sulfoxide (DMSO), was then added in the protein solution. The labeling reaction was kept at 4° C overnight, under gentle stirring. The unreacted spin label was removed using a PD-10 column and the purified fractions were checked acquiring X-band cw-EPR spectrum. The final protein concentration was estimated from the UV/vis spectrum at 280 nm by a Varian Cary 50 spectrophotometer (εUb = 1490 M-1 cm-1). In vitro samples

were inserted into EPR quartz tubes (1.6 mm O.D., 1 mm I.D.) and used to record cw-EPR spectra by a Bruker ELEXYS E580 spectrometer equipped with a super HIGH Q ER4122SHQE operating at X-band. Cw-EPR spectra were recorder with the following instrument settings: v= 9.874 GHz; center field= 3500 G; sweep width= 150 G; microwave power= 20 mW; modulation frequency= 100 kHz; modulation amplitude= 1 G; conversion time= 25 ms; sweep time= 25.6 s; scans= 25). All the in-cell EPR samples were prepared following the protocol reported in "heat shock insertion". Four consecutive wash steps with an isotonic PBS solution were performed and the surnatant of each step was checked acquiring the relative cw-EPR spectrum.

In-cell samples were transferred into EPR quartz tubes (1.6 mm O.D., 1 mm I.D.) and after a preparation time of ~10 minutes, RT cw-EPR spectra were recorded with the same spectrometer setup as above.

1.4. In cell sample preparation via heath shock delivery

E.coli

For E.coli cells we employed 50 µL of 500 µM doubly labelled human Ubiquitin (protonated or fully deuterated) solution in PBS buffer at pH 7.4 (Gibco®) to resuspend a pellet containing approximately 109 cells/mL E. coli DH5α competent cells. The cells were incubated with the protein solution at 42 °C for 1 min, using a thermic bath, to induce the internalization of the external protein. Afterwards the sample was left on ice for 2 minutes and was then centrifugate for 30 seconds. The supernatant was removed and the cells were then washed 4 times with an isotonic phosphate buffer solution to ensure complete removal of the protein which was not internalized into bacteria. Every wash step involved the addition of 200 µL of buffer, to completely resuspend the cell pellet, and a centrifugation step. The supernatant solution of each wash step was checked by EPR to assure the complete removal of the not-internalized spin labelled protein. After the last wash step the sample was resuspended in 50 µL of phosphate buffer and was directly inserted in Q-band tubes and snap freezed by complete immersion in liquid nitrogen. Every step was repeated for both protonated and fully deuterated bacterial cells.

It is worth mentioning that the whole thermal treatment needs to be carried out as fast as possible to avoid as much as possible the disappearance of the EPR nitroxide signal due to reduction phenomena that are present inside the intracellular environment. Previous works show how the complete disappearance of the signal can be observed after 40 minutes (ref).

Everything was repeated using both protonated and fully deuterated bacterial cells.

P.Pastoris

P. pastoris X-33 cells, previously stored in BEDS solution, were incubated at 30 °C for 10 min, using a thermoblock. The cells were then centrifuged at 2800 g for 3 min and then resuspended in 50 µL of TE buffer supplemented with 0.1% v/v lithium acetate. Then, 50 µL of 500 µM doubly labelled

human Ubiquitin (protonated or fully deuterated) solution was added to 2 x 108 cells/mL of P. pastoris X-33 cells. The yeast cells were finally incubated at 42 °C for 10 min to promote the protein internalization. The sample was then stored on ice for 3 minutes. The cells were then centrifuged, the protein solution was removed and the pellet was washed 4 times with isotonic PBS buffer. Every wash step involved the addition of 200 µL of buffer, to completely resuspend the cell pellet, and a centrifugation step. The supernatant solution of each wash step was checked by EPR to assure the complete removal of the not-internalized spin labelled protein. After the last wash step the sample was resuspended in 50 µL of phosphate buffer and was directly inserted in Q-band tubes and snap freezed by complete immersion in liquid nitrogen. Every step was repeated for both protonated and fully deuterated bacterial cells

1.5. X.-band Room Temperature measurements

In vitro and in cell samples were inserted into EPR quartz tubes (1.6 mm O.D., 1 mm I.D.) and used to record cw-EPR spectra by a Bruker ELEXYS E580 spectrometer equipped with a super HIGH Q ER4122SHQE operating at X-band. The in cell sample were inserted after after an avarege handling time of around 12-14 minutes. Cw-EPR spectra were recorder with the following instrument settings: v= 9.874 GHz; center field= 3500 G; sweep width= 150 G; microwave power= 20 mW; modulation frequency= 100 kHz; modulation amplitude= 1 G; conversion time= 25 ms; sweep time= 25.6 s; scans= 25). In-cell samples were transferred into EPR quartz tubes (1.6 mm O.D., 1 mm I.D.), In the case of in cell samples, in order to follow the bioreduction of the nitroxide spin label inside cells, a set cw-EPR experiments covering a time period of 120 minutes were recorded with the above spectroscopic parameters.

1.6. Q-band T_2 measurements

All the experiments were acquired using a Bruker ELEXYS E580 X/Q-band spectrometer equipped with a pulse 10 W Amp Q amplifier, an EN 5107D2 Q-band EPR/ENDOR probe-head and a continuous He-flow cryostat (Oxford Instruments) coupled with a temperature controller (Oxford Instruments). The temperature was kept at 50 K. All the T_2 experiments were acquired according to this pulse sequence: $\pi/2 - \tau - \pi - \tau$ under overcoupling condition. The experiments were acquired off resonance, 70 MHz from the resonator dip, to be consistent with the detection frequency in the DEER experiment. All pulses were optimized for each experiment using the regular mw-nutation sequence optimizing the pulse at the maximum of the nitroxide field. For all experiments the pulses were optimized through the use of the mw-nutation sequence.

2. Supplementary information

2.1. X-band Continuous Wave (CW) EPR

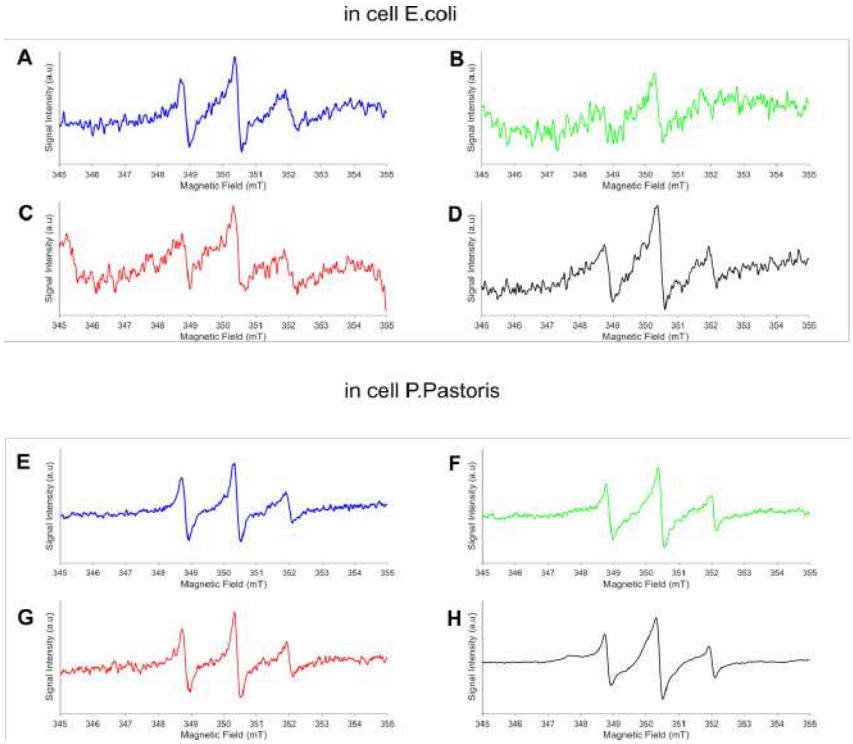


Figure S1: cw-EPR spectra acquired for the in cell sample prepared with *E. coli* and *P.pastoris* both in fully protonated and deuterated condition. In particular the blue experiments (A-E) were obtained by the in cell sample delivered using deuterated protein and deuterated cells. The red experiments (C-G) were obtained from the in cell sample prepared by delivering the deuterated protein inside the protonated cells. Green spectra (B-F) were measured from the in cell samples prepared delivering protonated protein inside protonated cells and, black experiments (D-H) were measured from the samples obtained by the delivery of protonated protein inside deuterated cells.

2.2. T_2 decays fitting for the in vitro and in cell samples

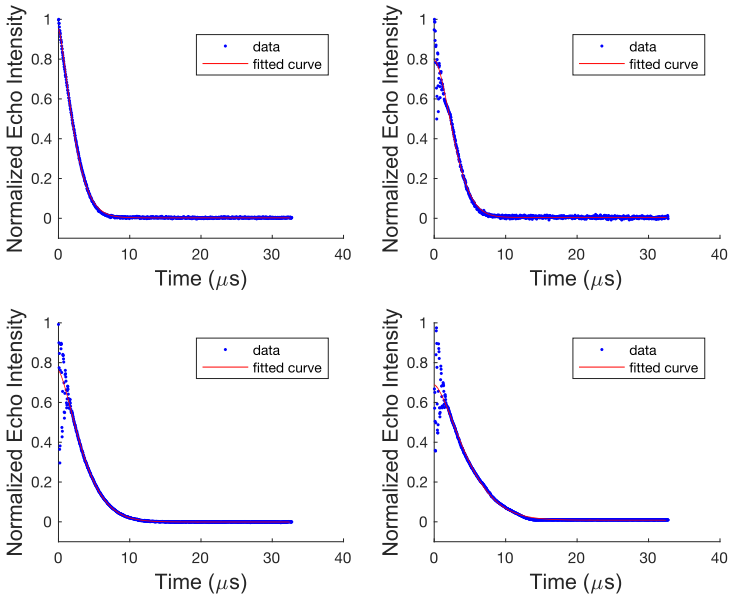


Figure S2: T_2 Fitting for the acquired in vitro samples. The acquired experimental data are represented by the blue dots, while the fitter curves are represented by the red lines. On the top left the in vitro sample prepared by placing the spin labelled protonated Ub in the protonated solvent , on the top left the in cell sample constituted by the protonated Ub resuspended in the deuterated buffer. On the bottom left, in cell sample prepared by having the deuterated Ub resuspended in the protonated buffer, and, on the bottom right the final sample constituted by the deuterated Ub in the deuterated buffer. All measurements were conducted as reported above and always in presence of 10 % v/v of glycerol (either deuterated or protonated for the relative buffer).

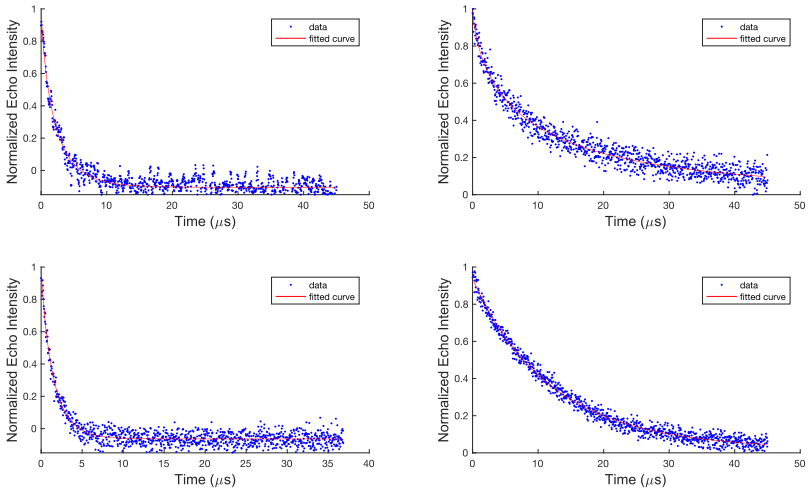


Figure S3: T_2 Fitting for the acquired in cell E.coli samples. The acquired experimental data are represented by the blue dots, while the fitter curves are represented by the red lines. On the top left the in cell sample prepared by delivering the protonated Ub inside the protonated cells , on the top right the in cell sample prepared by delivering the protonated Ub inside the deuterated cells. On the bottom left, in cell sample prepared by delivering the deuterated Ub inside the protonated cells , and , on the bottom right in cell sample prepared by delivering the deuterated Ub inside the deuterated cells.

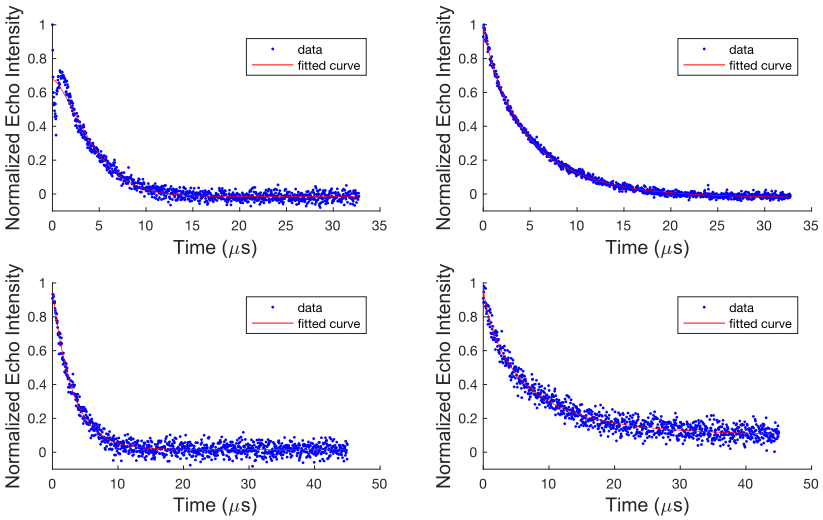


Figure S4: T_2 Fitting for the acquired in cell *P.pastoris* samples. The acquired experimental data are represented by the blue dots, while the fitter curves are represented by the red lines. On the top left the in cell sample prepared by delivering the protonated Ub inside the protonated cells, on the top left the in cell sample prepared by delivering the protonated Ub inside the deuterated cells. On the bottom left in cell sample prepared by delivering the deuterated Ub inside the protonated cells, and, on the bottom right in cell sample prepared by delivering the deuterated Ub inside the deuterated cells.

2.3. *In vitro T_2 concentration dependence*

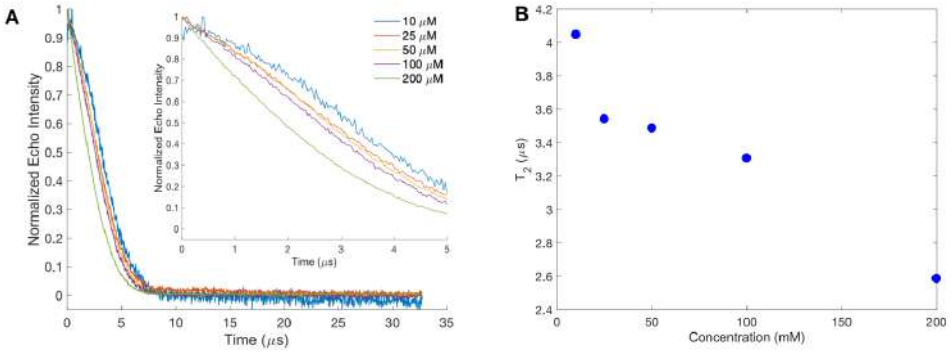


Figure S5: In vitro concentration dependent T_2 decays at Q-band frequency and field acquired at five different protein concentrations. The T_2 decays were acquired following the spectroscopical set-up reported in the Material and Methods section. As shown by the data the electronic T_2 increases its value by the increasing of the dilution of the spin labelled protein.

6. Conclusions

To conclude, this thesis work has explored different tag topologies: their characteristic and behaviour, the reaction needed to bioconjugate proteins and some of their possible applications to advance the field of structural biology. The use of ^{19}F-containing tags for post-expression covalent modification of tyrosine residues showcased a promising approach for enriching the toolkit available for protein analysis, especially in ^{19}F NMR studies. Additionally, our investigation into different nitroxide spin labels, strategically attached to cysteine residues on specific secondary structural elements (α-helix and β-sheet), provided critical insights into their behavior and effectiveness as probes. This exploration is crucial for optimizing tagging strategies and obtaining accurate structural information through techniques like EPR. By employing a Cu(II) chelate complex and leveraging pulse dipolar EPR techniques, we achieved precise distance measurements between orthogonal spin labels, advancing the application of EPR in analyzing protein structures. These findings contribute to the growing understanding of how specific tags and probes can enhance our ability to probe proteins at a molecular level. Moreover, the extension of our studies to in-cell environment, using various combinations of deuterated proteins and deuterated cells, allowed us to explore the impact of deuteration on the transverse relaxation time T_2. This investigation has implications on the optimization of experimental conditions and the improvement of the accuracy of distance measurements in cellular contexts, providing valuable information for researchers in the field of structural biology. Overall, this thesis contributes to enhance methodologies and strategies that are essential to study proteins and to unravel their complex biological roles.

7. References

1 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958). https://doi.org:10.1038/181662a0

2 Maier, O., Oberle, V. & Hoekstra, D. Fluorescent lipid probes: some properties and applications (a review). *Chem Phys Lipids* **116**, 3-18 (2002). https://doi.org:10.1016/s0009-3084(02)00017-8

3 Marsh, D. Electron spin resonance in membrane research: protein-lipid interactions. *Methods* **46**, 83-96 (2008). https://doi.org:10.1016/j.ymeth.2008.07.001

4 Torricella, F., Pierro, A., Mileo, E., Belle, V. & Bonucci, A. Nitroxide spin labels and EPR spectroscopy: A powerful association for protein dynamics studies. *Biochim Biophys Acta Proteins Proteom* **1869**, 140653 (2021). https://doi.org:10.1016/j.bbapap.2021.140653

5 Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J. & Waggoner, A. S. Cyanine Dye Labeling Reagents: Sulfoindocyanine Succinimidyl Esters. *Bioconjug Chem* **4**, 105-111 (1993). https://doi.org:10.1021/bc00020a001

6 Panchuk-Voloshina, N. *et al.* Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *J Histochem Cytochem* **47**, 1179-1188 (1999). https://doi.org:10.1177/002215549904700910

7 Schirripa Spagnolo, C. & Luin, S. Choosing the Probe for Single-Molecule Fluorescence Microscopy. *Int J Mol Sci* **23** (2022). https://doi.org:10.3390/ijms232314949

8 Nilsson, B. L., Soellner, M. B. & Raines, R. T. Chemical synthesis of proteins. *Annu Rev Biophys Biomol Struct* **34**, 91-118 (2005). https://doi.org:10.1146/annurev.biophys.34.040204.144700

9 Hemantha, H. P. *et al.* Nonenzymatic polyubiquitination of expressed proteins. *J Am Chem Soc* **136**, 2665-2673 (2014). https://doi.org:10.1021/ja412594d

10 Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. & Santi, D. V. Long-term stabilization of maleimide-thiol conjugates. *Bioconjug Chem* **26**, 145-152 (2015). https://doi.org:10.1021/bc5005262

11 Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* **48**, 6974-6998 (2009). https://doi.org:10.1002/anie.200900942

12 Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal Bioanal Chem* **392**, 305-312 (2008). https://doi.org:10.1007/s00216-008-2231-5

13 Hermanson, G. T. *Bioconjugate Techniques*. (2nd edn Academic Press, Inc., 2008).

14 Nakamura, T., Kawai, Y., Kitamoto, N., Osawa, T. & Kato, Y. Covalent modification of lysine residues by allyl isothiocyanate in physiological conditions: plausible transformation of isothiocyanate from thiol to amine. *Chem Res Toxicol* **22**, 536-542 (2009). https://doi.org:10.1021/tx8003906

15 Jentoft, N. & Dearborn, D. G. Labeling of proteins by reductive methylation using sodium cyanoborohydride. *J Biol Chem* **254**, 4359-4365 (1979).

16 Hooker, J. M., Kovacs, E. W. & Francis, M. B. Interior surface modification of bacteriophage MS2. *J Am Chem Soc* **126**, 3718-3719 (2004). https://doi.org:10.1021/ja0317909

17 Schlick, T. L., Ding, Z., Kovacs, E. W. & Francis, M. B. Dual-surface modification of the tobacco mosaic virus. *J Am Chem Soc* **127**, 3718-3723 (2005). https://doi.org:10.1021/ja046239n

18 Sato, S. *et al.* Site-Selective Protein Chemical Modification of Exposed Tyrosine Residues Using Tyrosine Click Reaction. *Bioconjug Chem* **31**, 1417-1424 (2020). https://doi.org:10.1021/acs.bioconjchem.0c00120

19 Joshi, N. S., Whitaker, L. R. & Francis, M. B. A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J Am Chem Soc* **126**, 15942-15943 (2004). https://doi.org:10.1021/ja0439017

20 Jia, S., He, D. & Chang, C. J. Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation. *J Am Chem Soc* **141**, 7294-7301 (2019). https://doi.org:10.1021/jacs.8b11912

21 Koniev, O. & Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem Soc Rev* **44**, 5495-5551 (2015). https://doi.org:10.1039/c5cs00048c

22 Voss, J., Salwinski, L., Kaback, H. R. & Hubbell, W. L. A method for distance determination in proteins using a designed metal ion binding site and site-directed spin labeling: evaluation with T4 lysozyme. *Proc Natl Acad Sci U S A* **92**, 12295-12299 (1995). https://doi.org:10.1073/pnas.92.26.12295

23 Cunningham, T. F., Putterman, M. R., Desai, A., Horne, W. S. & Saxena, S. The double-histidine Cu^{2+}-binding motif: a highly rigid, site-specific spin probe for electron spin resonance distance measurements. *Angew Chem Int Ed Engl* **54**, 6330-6334 (2015). https://doi.org:10.1002/anie.201501968

24 Chen, H., Viel, S., Ziarelli, F. & Peng, L. ^{19}F NMR: a valuable tool for studying biological events. *Chem Soc Rev* **42**, 7971-7982 (2013). https://doi.org:10.1039/c3cs60129c

25 Danielson, M. A. & Falke, J. J. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct* **25**, 163-195 (1996). https://doi.org:10.1146/annurev.bb.25.060196.001115

26 Sharaf, N. G. & Gronenborn, A. M. (19)F-modified proteins and (19)F-containing ligands as tools in solution NMR studies of protein interactions. *Methods Enzymol* **565**, 67-95 (2015). https://doi.org:10.1016/bs.mie.2015.05.014

27 Quint, P. *et al.* Structural mobility in human manganese superoxide dismutase. *Biochemistry* **45**, 8209-8215 (2006). https://doi.org:10.1021/bi060628s

28 Papeo, G. *et al.* Polyfluorinated amino acids for sensitive ^{19}F NMR-based screening and kinetic measurements. *J Am Chem Soc* **129**, 5665-5672 (2007). https://doi.org:10.1021/ja069128s

29 Yu, L., Hajduk, P. J., Mack, J. & Olejniczak, E. T. Structural studies of Bcl-xL/ligand complexes using ^{19}F NMR. *J Biomol NMR* **34**, 221-227 (2006). https://doi.org:10.1007/s10858-006-0005-y

30 Bann, J. G., Pinkner, J., Hultgren, S. J. & Frieden, C. Real-time and equilibrium ^{19}F-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD. *Proc Natl Acad Sci U S A* **99**, 709-714 (2002). https://doi.org:10.1073/pnas.022649599

31 Li, H. & Frieden, C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and ^{19}F NMR. *Proc Natl Acad Sci U S A* **104**, 11993-11998 (2007). https://doi.org:10.1073/pnas.0705253104

32 Pham, L. B. T. *et al.* Direct Expression of Fluorinated Proteins in Human Cells for ^{19}F In-Cell NMR Spectroscopy. *J Am Chem Soc* **145**, 1389-1399 (2023). https://doi.org:10.1021/jacs.2c12086

33 Zhu, W. *et al.* Visualizing Proteins in Mammalian Cells by ^{19} F NMR Spectroscopy. *Angew Chem Int Ed Engl* **61**, e202201097 (2022). https://doi.org:10.1002/anie.202201097

34 Li, C. *et al.* Protein ^{19}F NMR in *Escherichia coli*. *J Am Chem Soc* **132**, 321-327 (2010). https://doi.org:10.1021/ja907966n

35 Salwiczek, M., Nyakatura, E. K., Gerling, U. I. M., Ye, S. J. & Koksch, B. Fluorinated amino acids: compatibility with native protein structures and effects on protein-protein interactions. *Chemical Society Reviews* **41**, 2135-2171 (2012). https://doi.org:10.1039/c1cs15241f

36 Welte, H., Zhou, T. K., Mihajlenko, X., Mayans, O. & Kovermann, M. What does fluorine do to a protein? Thermodynamic, and highly-resolved structural insights into fluorine-labelled variants of the cold shock protein. *Sci Rep-Uk* **10** (2020). https://doi.org:ARTN 2640
10.1038/s41598-020-59446-w

37 Kim, H. W., Perez, J. A., Ferguson, S. J. & Campbell, I. D. The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. Application to fluorotryptophan labelling to the H(+)-ATPase of Escherichia coli and NMR studies. *FEBS Lett* **272**, 34-36 (1990). https://doi.org:10.1016/0014-5793(90)80442-l

38 Gimenez, D. A.-O., Phelan, A. A.-O., Murphy, C. A.-O. & Cobb, S. A.-O. ^{19}F NMR as a tool in chemical biology. *Beilstein J Org Chem* **17**, 293-318 (2021). https://doi.org:10.3762/bjoc.17.28

39 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001). https://doi.org:10.1126/science.1060077

40 Gronenborn, A. M. Small, but powerful and attractive: ^{19}F in biomolecular NMR. *Structure* **30**, 6-14 (2022). https://doi.org:10.1016/j.str.2021.09.009

41 Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of ^{19}F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* **62**, 1-33 (2012). https://doi.org:10.1016/j.pnmrs.2011.06.003

42 Luchette, P. A., Prosser, R. S. & Sanders, C. R. Oxygen as a Paramagnetic Probe of Membrane Protein Structure by Cysteine Mutagenesis and ^{19}F NMR Spectroscopy. *J Am Chem Soc* **124**, 1778-1781 (2002). https://doi.org:10.1021/ja016748e

43 Thomas, M. R. & Boxer, S. G. 19F NMR of trifluoroacetyl-labeled cysteine mutants of myoglobin: structural probes of nitric oxide bound to the H93G cavity mutant. *Biochemistry* **40**, 8588-8596 (2001). https://doi.org:10.1021/bi010187

44 Hellmich, U. A., Pfleger, N. & Glaubitz, C. F-MAS NMR on proteorhodopsin: enhanced protocol for site-specific labeling for general application to membrane proteins. *Photochem Photobiol* **85**, 535-539 (2009). https://doi.org:10.1111/j.1751-1097.2008.00498.x

45 Manglik, A. *et al.* Structural insights into the dynamic process of β_2-adrenergic receptor signaling. *Cell* **161**, 1101-1111 (2015). https://doi.org:10.1016/j.cell.2015.04.043

46 Edwards, J. M., Derrick, J. P., van der Walle, C. F. & Golovanov, A. P. ^{19}F NMR as a Tool for Monitoring Individual Differentially Labeled Proteins in Complex Mixtures. *Mol Pharm* **15**, 2785-2796 (2018). https://doi.org:10.1021/acs.molpharmaceut.8b00282

47 Klein-Seetharaman, J., Getmanova, E. V., Loewen, M. C., Reeves, P. J. & Khorana, H. G. NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: applicability of solution (19)F NMR. *Proc Natl Acad Sci U S A* **96**, 13744-13749 (1999). https://doi.org:10.1073/pnas.96.24.13744

48 Susac, L., Eddy, M. T., Didenko, T., Stevens, R. C. & Wuthrich, K. A_{2A} adenosine receptor functional states characterized by ^{19}F-NMR. *Proc Natl Acad Sci U S A* **115**, 12733-12738 (2018). https://doi.org:10.1073/pnas.1813649115

49 Kalbitzer, H. R. *et al.* A new high sensitivity 19F probe for labeling cysteine groups of proteins. *NMR Biomed* **5**, 347-350 (1992). https://doi.org:10.1002/nbm.1940050605

50 Larda, S. T., Pichugin, D. & Prosser, R. S. Site-Specific Labeling of Protein Lysine Residues and N-Terminal Amino Groups with Indoles and Indole-Derivatives. *Bioconjug Chem* **26**, 2376-2383 (2015). https://doi.org:10.1021/acs.bioconjchem.5b00457

51 Higgins, H. G. & Harrington, K. J. Reaction of amino acids and proteins with diazonium compounds. II. Spectra of protein derivatives. *Arch Biochem Biophys* **85**, 409-425 (1959). https://doi.org:10.1016/0003-9861(59)90506-5

52 Jones, M. W. *et al.* Direct peptide bioconjugation/PEGylation at tyrosine with linear and branched polymeric diazonium salts. *J Am Chem Soc* **134**, 7406-7413 (2012). https://doi.org:10.1021/ja2118559

53 Gavrilyuk, J., Ban, H., Nagano, M., Hakamata, W. & Barbas, C. F., 3rd. Formylbenzene diazonium hexafluorophosphate reagent for tyrosine-selective modification of proteins and the introduction of a bioorthogonal aldehyde. *Bioconjug Chem* **23**, 2321-2328 (2012). https://doi.org:10.1021/bc300410p

54 Ban, H. *et al.* Facile and stabile linkages through tyrosine: bioconjugation strategies with the tyrosine-click reaction. *Bioconjug Chem* **24**, 520-532 (2013). https://doi.org:10.1021/bc300665t

55 Ban, H., Gavrilyuk, J. & Barbas, C. F., 3rd. Tyrosine bioconjugation through aqueous ene-type reactions: a click-like reaction for tyrosine. *J Am Chem Soc* **132**, 1523-1525 (2010). https://doi.org:10.1021/ja9090624

56 Hu, Q.-Y. *et al.* Synthesis of a well-defined glycoconjugate vaccine by a tyrosine-selective conjugation strategy. *Chemical Science* **4**, 3827-3832 (2013). https://doi.org:10.1039/C3SC51694F

57 Romanini, D. W. & Francis, M. B. Attachment of peptide building blocks to proteins through tyrosine bioconjugation. *Bioconjug Chem* **19**, 153-157 (2008). https://doi.org:10.1021/bc700231v

58 Mileo, E. *et al.* Enlarging the panoply of site-directed spin labeling electron paramagnetic resonance (SDSL-EPR): sensitive and selective spin-labeling of tyrosine using an isoindoline-based nitroxide. *Bioconjug Chem* **24**, 1110-1117 (2013). https://doi.org:10.1021/bc4000542

59 Commoner, B., Townsend, J. & Pake, G. E. Free radicals in biological materials. *Nature* **174**, 689-691 (1954). https://doi.org:10.1038/174689a0

60 Gordy, W., Ard, W. B. & Shields, H. Microwave Spectroscopy of Biological Substances. I. Paramagnetic Resonance in X-Irradiated Amino Acids and Proteins. *Proc Natl Acad Sci U S A* **41**, 983-996 (1955). https://doi.org:10.1073/pnas.41.11.983

61 Windle, J. J., Wiersema, A. K., Clark, J. R. & Feeney, R. E. Investigation of the Iron and Copper Complexes of Avian Conalbumins and Human Transferrins by Electron Paramagnetic Resonance. *Biochemistry* **2**, 1341-1345 (1963). https://doi.org:10.1021/bi00906a028

62 Berliner, L. J., Grunwald, J., Hankovszky, H. O. & Hideg, K. A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Anal Biochem* **119**, 450-455 (1982). https://doi.org:10.1016/0003-2697(82)90612-1

63 Kurshev, V. V., Raitsimring, A. M. & Tsvetkov, Y. D. Selection of dipolar interaction by the "2 + 1" pulse train ESE. *Journal of Magnetic Resonance (1969)* **81**, 441-454 (1989). https://doi.org:https://doi.org/10.1016/0022-2364(89)90080-2

64 Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science* **248**, 1088-1092 (1990). https://doi.org:10.1126/science.2160734

65 Altenbach, C., Flitsch, S. L., Khorana, H. G. & Hubbell, W. L. Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. *Biochemistry* **28**, 7806-7812 (1989). https://doi.org:10.1021/bi00445a042

66 Altenbach, C., Froncisz, W., Hyde, J. S. & Hubbell, W. L. Conformation of spin-labeled melittin at membrane surfaces investigated by pulse saturation recovery and continuous wave power saturation electron paramagnetic resonance. *Biophys J* **56**, 1183-1191 (1989). https://doi.org:10.1016/S0006-3495(89)82765-1

67 Todd, A. P., Cong, J., Levinthal, F., Levinthal, C. & Hubbell, W. L. Site-directed mutagenesis of colicin E1 provides specific attachment sites for spin labels whose spectra are sensitive to local conformation. *Proteins* **6**, 294-305 (1989). https://doi.org:10.1002/prot.340060312

68 Jeschke, G., Pannier, M., Godt, A. & Spiess, H. W. Dipolar spectroscopy and spin alignment in electron paramagnetic resonance. *Chemical Physics Letters* **331**, 243-252 (2000). https://doi.org:https://doi.org/10.1016/S0009-2614(00)01171-4

69 Borbat, P. P. & Freed, J. H. Multiple-quantum ESR and distance measurements. *Chemical Physics Letters* **313**, 145-154 (1999). https://doi.org:https://doi.org/10.1016/S0009-2614(99)00972-0

70 Martin, R. E. *et al.* Determination of End-to-End Distances in a Series of TEMPO Diradicals of up to 2.8 nm Length with a New Four-Pulse Double Electron Electron Resonance Experiment. *Angew Chem Int Ed Engl* **37**, 2833-2837 (1998). https://doi.org:10.1002/(SICI)1521-3773(19981102)37:20<2833::AID-ANIE2833>3.0.CO;2-7

71 Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *J Magn Reson* **142(2)**, 331-340 (2000). https://doi.org:https://doi.org/10.1006/jmre.1999.1944

72 Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J Magn Reson* **172**, 279-295 (2005). https://doi.org:10.1016/j.jmr.2004.10.012

73 Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Applied Magnetic Resonance* **30**, 473-498 (2006). https://doi.org:10.1007/BF03166213

74 Jeschke, G. *et al.* Localization of the N-terminal Domain in Light-harvesting Chlorophyll *a*/*b* Protein by EPR Measurements. *Journal of Biological Chemistry* **280**, 18623-18630 (2005). https://doi.org:https://doi.org/10.1074/jbc.M501171200

75 Ahammad, T. *et al.* Continuous Wave Electron Paramagnetic Resonance Spectroscopy Reveals the Structural Topology and Dynamic Properties of Active Pinholin $S^{21}68$ in a Lipid Bilayer. *J Phys Chem B* **123**, 8048-8056 (2019). https://doi.org:10.1021/acs.jpcb.9b06480

76 Galazzo, L. *et al.* Identifying conformational changes with site-directed spin labeling reveals that the GTPase domain of HydF is a molecular switch. *Sci Rep* **7**, 1714 (2017). https://doi.org:10.1038/s41598-017-01886-y

77 Lindemann, W. R. *et al.* Conformational Dynamics in Extended RGD-Containing Peptides. *Biomacromolecules* **21**, 2786-2794 (2020). https://doi.org:10.1021/acs.biomac.0c00506

78 Yardeni, E. H. *et al.* Probing the solution structure of the *E. coli* multidrug transporter MdfA using DEER distance measurements with nitroxide and Gd(III) spin labels. *Sci Rep* **9**, 12528 (2019). https://doi.org:10.1038/s41598-019-48694-0

79 Strohaker, T. *et al.* Structural heterogeneity of α-synuclein fibrils amplified from patient brain extracts. *Nat Commun* **10**, 5535 (2019). https://doi.org:10.1038/s41467-019-13564-w

80 Chuo, S. W. *et al.* Conformational Response of N-Terminally Truncated Cytochrome P450 3A4 to Ligand Binding in Solution. *Biochemistry* **58**, 3903-3910 (2019). https://doi.org:10.1021/acs.biochem.9b00620

81 Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**, 596-599 (2010). https://doi.org:10.1038/nature09454

82 Schiemann, O. *et al.* A PELDOR-based Nanometer Distance Ruler for Oligonucleotides. *J Am Chem Soc* **126**, 5722-5729 (2004). https://doi.org:10.1021/ja0393877

83 Wuebben, C. *et al.* Site-Directed Spin Labeling of RNA with a *Gem*-Diethylisoindoline Spin Label: PELDOR, Relaxation, and Reduction Stability. *Molecules* **24** (2019). https://doi.org:10.3390/molecules24244482

84 Duss, O., Yulikov, M., Jeschke, G. & Allain, F. H. EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. *Nat Commun* **5**, 3669 (2014). https://doi.org:10.1038/ncomms4669

85 Dalaloyan, A. *et al.* Tracking Conformational Changes in Calmodulin in vitro, in Cell Extract, and in Cells by Electron Paramagnetic Resonance Distance Measurements. *Chemphyschem* **20**, 1860-1868 (2019). https://doi.org:10.1002/cphc.201900341

86 Joseph, B., Sikora, A. & Cafiso, D. S. Ligand Induced Conformational Changes of a Membrane Transporter in *E. coli* Cells Observed with DEER/PELDOR. *J Am Chem Soc* **138**, 1844-1847 (2016). https://doi.org:10.1021/jacs.5b13382

87 Kapsalis, C. *et al.* Allosteric activation of an ion channel triggered by modification of mechanosensitive nano-pockets. *Nat Commun* **10**, 4619 (2019). https://doi.org:10.1038/s41467-019-12591-x

88 Hutter, C. A. J. *et al.* The extracellular gate shapes the energy profile of an ABC exporter. *Nat Commun* **10**, 2260 (2019). https://doi.org:10.1038/s41467-019-09892-6

89 Yang, Y., Yang, F., Li, X. Y., Su, X. C. & Goldfarb, D. In-Cell EPR Distance Measurements on Ubiquitin Labeled with a Rigid PyMTA-Gd(III) Tag. *J Phys Chem B* **123**, 1050-1059 (2019). https://doi.org:10.1021/acs.jpcb.8b11442

90 Yang, Y. *et al.* In-cell destabilization of a homodimeric protein complex detected by DEER spectroscopy. *Proc Natl Acad Sci U S A* **117**, 20566-20575 (2020). https://doi.org:10.1073/pnas.2005779117

91 Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson* **178**, 42-55 (2006). https://doi.org:10.1016/j.jmr.2005.08.013

92 Schweiger, A. & Jeschke, G. *Principles of Pulse Electron Paramagnetic Resonance*. (Oxford Univ. Press, 2001).

93 Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance* (2002).

94 Robinson, J. W. *Pratical Handbook of Spectroscopy*. (Routledge, 2017).

95 Milov, A. D., Maryasov, A. G. & Tsvetkov, Y. D. Pulsed electron double resonance (PELDOR) and its applications in free-radicals research. *Applied Magnetic Resonance* **15**, 107-143 (1998). https://doi.org:10.1007/BF03161886

96 Prisner, T., Rohrer, M. & MacMillan, F. Pulsed EPR spectroscopy: biological applications. *Annu Rev Phys Chem* **52**, 279-313 (2001). https://doi.org:10.1146/annurev.physchem.52.1.279

97 Goldfarb, D. Pulse EPR in biological systems - Beyond the expert's courtyard. *J Magn Reson* **306**, 102-108 (2019). https://doi.org:10.1016/j.jmr.2019.07.038

98 Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* **9**, 1895-1910 (2007). https://doi.org:10.1039/b614920k

99 Schmidt, T., Walti, M. A., Baber, J. L., Hustedt, E. J. & Clore, G. M. Long Distance Measurements up to 160 Å in the GroEL Tetradecamer Using Q-Band DEER EPR Spectroscopy. *Angew Chem Int Ed Engl* **55**, 15905-15909 (2016). https://doi.org:10.1002/anie.201609617

100 Milov, A. D., Salikhov, K. M. & M.D., S. Application of the double resonance method to electron spin echo in a study of the spatial distribution of paramagnetic centers in solids. *Sov. Phys. Solid State*, 565-569 (1981).

101 Milikisyants, S., Scarpelli, F., Finiguerra, M. G., Ubbink, M. & Huber, M. A pulsed EPR method to determine distances between paramagnetic centers with strong spectral anisotropy and radicals: the dead-time free RIDME sequence. *J Magn Reson* **201**, 48-56 (2009). https://doi.org:10.1016/j.jmr.2009.08.008

102 Kulik, L. V., Dzuba, S. A., Grigoryev, I. A. & Tsvetkov, Y. D. Electron dipole–dipole interaction in ESEEM of nitroxide biradicals. *Chemical Physics Letters* **343**, 315-324 (2001). https://doi.org:https://doi.org/10.1016/S0009-2614(01)00721-7

103 Wort, J. L., Ackermann, K., Giannoulis, A. & Bode, B. E. Enhanced sensitivity for pulse dipolar EPR spectroscopy using variable-time RIDME. *J Magn Reson* **352**, 107460 (2023). https://doi.org:10.1016/j.jmr.2023.107460

104 Giannoulis, A., Ben-Ishay, Y. & Goldfarb, D. Characteristics of Gd(III) spin labels for the study of protein conformations. *Methods Enzymol* **651**, 235-290 (2021). https://doi.org:10.1016/bs.mie.2021.01.040

105 Wort, J. L. *et al.* Sub-Micromolar Pulse Dipolar EPR Spectroscopy Reveals Increasing Cu^{II}-labelling of Double-Histidine Motifs with Lower Temperature. *Angew Chem Int Ed Engl* **58**, 11681-11685 (2019). https://doi.org:10.1002/anie.201904848

106 Schiemann, O. & Prisner, T. F. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q Rev Biophys* **40**, 1-53 (2007). https://doi.org:10.1017/S003358350700460X

107 Giannoulis, A. *et al.* Two closed ATP- and ADP-dependent conformations in yeast Hsp90 chaperone detected by Mn(II) EPR spectroscopic techniques. *Proc Natl Acad Sci U S A* **117**, 395-404 (2020). https://doi.org:10.1073/pnas.1916030116

108 Kuzhelev, A. A. *et al.* Triarylmethyl Radicals: EPR Study of ^{13}C Hyperfine Coupling Constants. *Z Phys Chem (N F)* **231**, 777-794 (2017). https://doi.org:10.1515/zpch-2016-0811

109 Fleck, N. *et al.* SLIM: A Short-Linked, Highly Redox-Stable Trityl Label for High-Sensitivity In-Cell EPR Distance Measurements. *Angew Chem Int Ed Engl* **59**, 9767-9772 (2020). https://doi.org:10.1002/anie.202004452

110 Toniolo, C. *et al.* Synthesis and conformational studies of peptides containing TOAC, a spin-labelled $C^{\alpha,\alpha}$-disubstituted glycine. *J Pept Sci* **1**, 45-57 (1995). https://doi.org:10.1002/psc.310010107

111 Jeschke, G. DEER distance measurements on proteins. *Annu Rev Phys Chem* **63**, 419-446 (2012). https://doi.org:10.1146/annurev-physchem-032511-143716

112 Pali, T., Finbow, M. E. & Marsh, D. Membrane assembly of the 16-kDa proteolipid channel from *Nephrops norvegicus* studied by relaxation enhancements in spin-label ESR. *Biochemistry* **38**, 14311-14319 (1999). https://doi.org:10.1021/bi991459c

113 Shenberger, Y., Gottlieb, H. E. & Ruthstein, S. EPR and NMR spectroscopies provide input on the coordination of Cu(I) and Ag(I) to a disordered methionine segment. *J Biol Inorg Chem* **20**, 719-727 (2015). https://doi.org:10.1007/s00775-015-1259-1

114 Dugas, H., Rodriguez, A. & Brisson, N. Selective spin-labeling of the ribosomal proteins of 70S ribosomes from Escherichia coli. *Can J Biochem.* **57**, 1407-1415 (1979). https://doi.org:10.1139/079-187

115 Gmeiner, C., Dorn, G., Allain, F. H. T., Jeschke, G. & Yulikov, M. Spin labelling for integrative structure modelling: a case study of the polypyrimidine-tract binding protein 1 domains in complexes with short RNAs. *Phys Chem Chem Phys* **19**, 28360-28380 (2017). https://doi.org:10.1039/c7cp05822e

116 Volkov, A., Dockter, C., Bund, T., Paulsen, H. & Jeschke, G. Pulsed EPR determination of water accessibility to spin-labeled amino acid residues in LHCIIb. *Biophys J* **96**, 1124-1141 (2009). https://doi.org:10.1016/j.bpj.2008.09.047

117 Hewgley, P. B. & Puett, D. Spin labeled calmodulin: a new probe for studying Ca^{2+} and macromolecular interactions. *Ann N Y Acad Sci* **356**, 20-32 (1980). https://doi.org:10.1111/j.1749-6632.1980.tb29595.x

118 Balo, A. R., Feyrer, H. & Ernst, O. P. Toward Precise Interpretation of DEER-Based Distance Distributions: Insights from Structural Characterization of V1 Spin-Labeled Side Chains. *Biochemistry* **55**, 5256-5263 (2016). https://doi.org:10.1021/acs.biochem.6b00608

119 Wingler, L. M. *et al.* Angiotensin Analogs with Divergent Bias Stabilize Distinct Receptor Conformations. *Cell* **176**, 468-478 e411 (2019). https://doi.org:10.1016/j.cell.2018.12.005

120 Toledo Warshaviak, D., Khramtsov, V. V., Cascio, D., Altenbach, C. & Hubbell, W. L. Structure and dynamics of an imidazoline nitroxide side chain with strongly hindered internal motion in proteins. *J Magn Reson* **232**, 53-61 (2013). https://doi.org:10.1016/j.jmr.2013.04.013

121 Khramtsov, V. V. *et al.* Quantitative determination of SH groups in low- and high-molecular-weight compounds by an electron spin resonance method. *Anal Biochem* **182**, 58-63 (1989). https://doi.org:10.1016/0003-2697(89)90718-5

122 Ackermann, K., Chapman, A. & Bode, B. E. A Comparison of Cysteine-Conjugated Nitroxide Spin Labels for Pulse Dipolar EPR Spectroscopy. *Molecules* **26** (2021). https://doi.org:10.3390/molecules26247534

123 Ji, M., Ruthstein, S. & Saxena, S. Paramagnetic metal ions in pulsed ESR distance distribution measurements. *Acc Chem Res* **47**, 688-695 (2014). https://doi.org:10.1021/ar4002452

124 Merz, G. E. *et al.* Copper-based pulsed dipolar ESR spectroscopy as a probe of protein conformation linked to disease states. *Biophys J* **107**, 1669-1674 (2014). https://doi.org:10.1016/j.bpj.2014.07.068

125 van Wonderen, J. H., Kostrz, D. N., Dennison, C. & MacMillan, F. Refined distances between paramagnetic centers of a multi-copper nitrite reductase determined by pulsed EPR (*i*DEER) spectroscopy. *Angew Chem Int Ed Engl* **52**, 1990-1993 (2013). https://doi.org:10.1002/anie.201208166

126 Cunningham, T. F. *et al.* Cysteine-specific Cu2+ chelating tags used as paramagnetic probes in double electron electron resonance. *J Phys Chem B* **119**, 2839-2843 (2015). https://doi.org:10.1021/jp5103143

127 Lawless, M. J., Ghosh, S., Cunningham, T. F., Shimshi, A. & Saxena, S. On the use of the Cu^{2+}-iminodiacetic acid complex for double histidine based distance measurements by pulsed ESR. *Phys Chem Chem Phys* **19**, 20959-20967 (2017). https://doi.org:10.1039/c7cp02564e

128 Ghosh, S., Lawless, M. J., Rule, G. S. & Saxena, S. The Cu^{2+}-nitrilotriacetic acid complex improves loading of alpha-helical double histidine site for precise distance measurements by pulsed ESR. *J Magn Reson* **286**, 163-171 (2018). https://doi.org:10.1016/j.jmr.2017.12.005

129 Sarver, J. L., Townsend, J. E., Rajapakse, G., Jen-Jacobson, L. & Saxena, S. Simulating the dynamics and orientations of spin-labeled side chains in a protein-DNA complex. *J Phys Chem B* **116**, 4024-4033 (2012). https://doi.org:10.1021/jp211094n

130 Hagelueken, G., Ward, R., Naismith, J. H. & Schiemann, O. MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Appl Magn Reson* **42**, 377-391 (2012). https://doi.org:10.1007/s00723-012-0314-0

131 Jeschke, G. MMM: A toolbox for integrative structure modeling. *Protein Sci* **27**, 76-85 (2018). https://doi.org:10.1002/pro.3269

132 Ghosh, S., Saxena, S. & Jeschke, G. Rotamer Modelling of Cu(II) Spin Labels Based on the Double-Histidine Motif. *Applied Magnetic Resonance* **49**, 1281-1298 (2018). https://doi.org:10.1007/s00723-018-1052-8

133 Vitali, V., Ackermann, K., Hagelueken, G. & Bode, B. E. Spectroscopically orthogonal labelling to disentangle site-specific nitroxide label distributions (Dataset). *In Dataset* **University of St Andrews Research Portal** (2023). https://doi.org:https://doi.org/10.17630/71f8e2e5-9f57-4160-8c32-de1b37d4c073

134 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

135 Wu, R. *et al.* High-resolution de novo structure prediction from primary sequence. *bioRxiv*, 2022.2007.2021.500999 (2022). https://doi.org:10.1101/2022.07.21.500999

136 Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023). https://doi.org:10.1126/science.ade2574

137 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021). https://doi.org:10.1126/science.abj8754

138 Luchinat, E. & Banci, L. A Unique Tool for Cellular Structural Biology: In-cell NMR. *J Biol Chem* **291**, 3776-3784 (2016). https://doi.org:10.1074/jbc.R115.643247

139 Bonucci, A., Ouari, O., Guigliarelli, B., Belle, V. & Mileo, E. In-Cell EPR: Progress towards Structural Studies Inside Cells. *Chembiochem* **21**, 451-460 (2020). https://doi.org:10.1002/cbic.201900291

140 Serber, Z. *et al.* High-resolution macromolecular NMR spectroscopy inside living cells. *J Am Chem Soc* **123**, 2446-2447 (2001). https://doi.org:10.1021/ja0057528

141 Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458**, 106-109 (2009). https://doi.org:10.1038/nature07839

142 Theillet, F. X. In-Cell Structural Biology by NMR: The Benefits of the Atomic Scale. *Chem Rev* **122**, 9497-9570 (2022). https://doi.org:10.1021/acs.chemrev.1c00937

143 Goldfarb, D. Exploring protein conformations in vitro and in cell with EPR distance measurements. *Curr Opin Struct Biol* **75**, 102398 (2022). https://doi.org:10.1016/j.sbi.2022.102398

144 Hamatsu, J. *et al.* High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J Am Chem Soc* **135**, 1688-1691 (2013). https://doi.org:10.1021/ja310928u

145 Banci, L., Barbieri, L., Bertini, I., Cantini, F. & Luchinat, E. In-cell NMR in E. coli to monitor maturation steps of hSOD1. *PLoS One* **6**, e23561 (2011). https://doi.org:10.1371/journal.pone.0023561

146 Crowley, P. B., Chow, E. & Papkovskaia, T. Protein interactions in the Escherichia coli cytosol: an impediment to in-cell NMR spectroscopy. *Chembiochem* **12**, 1043-1048 (2011). https://doi.org:10.1002/cbic.201100063

147 Torricella, F., Bonucci, A., Polykretis, P., Cencetti, F. & Banci, L. Rapid protein delivery to living cells for biomolecular investigation. *Biochem Biophys Res Commun* **570**, 82-88 (2021). https://doi.org:10.1016/j.bbrc.2021.07.006

148 Wall, K. P. & Hough, L. E. In-Cell NMR within Budding Yeast Reveals Cytoplasmic Masking of Hydrophobic Residues of FG Repeats. *Biophys J* **115**, 1690-1695 (2018). https://doi.org:10.1016/j.bpj.2018.08.049

149 Luchinat, E. *et al.* Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency. *Angew Chem Int Ed Engl* **59**, 6535-6539 (2020). https://doi.org:10.1002/anie.201913436

150 Polykretis, P., Cencetti, F., Donati, C., Luchinat, E. & Banci, L. Cadmium effects on superoxide dismutase 1 in human cells revealed by NMR. *Redox Biol* **21**, 101102 (2019). https://doi.org:10.1016/j.redox.2019.101102

151 Torricella, F., Barbieri, L., Bazzurro, V., Diaspro, A. & Banci, L. Protein delivery to living cells by thermal stimulation for biophysical investigation. *Sci Rep* **12**, 17190 (2022). https://doi.org:10.1038/s41598-022-21103-9

152 Hubbell, W. L. & McConnell, H. M. Molecular motion in spin-labeled phospholipids and membranes. *J Am Chem Soc* **93**, 314-326 (1971). https://doi.org:10.1021/ja00731a005

153 Igarashi, R. *et al.* Distance determination in proteins inside Xenopus laevis oocytes by double electron-electron resonance experiments. *J Am Chem Soc* **132**, 8228-8229 (2010). https://doi.org:10.1021/ja906104e

154 Hasanbasri, Z., Singewald, K., Gluth, T. D., Driesschaert, B. & Saxena, S. Cleavage-Resistant Protein Labeling With Hydrophilic Trityl Enables Distance Measurements *In-Cell*. *J Phys Chem B* **125**, 5265-5274 (2021). https://doi.org:10.1021/acs.jpcb.1c02371

155 Collauto, A. *et al.* Compaction of RNA Duplexes in the Cell. *Angew Chem Int Ed Engl* **59**, 23025-23029 (2020). https://doi.org:10.1002/anie.202009800

156 Ovcherenko, S. S. *et al.* Uptake of Cell-Penetrating Peptide RL2 by Human Lung Cancer Cells: Monitoring by Electron Paramagnetic Resonance and Confocal Laser Scanning Microscopy. *Molecules* **26** (2021). https://doi.org:10.3390/molecules26185442

157 Ogino, S. *et al.* Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J Am Chem Soc* **131**, 10834-10835 (2009). https://doi.org:10.1021/ja904407w

158 Dunkel, S., Pulagam, L. P., Steinhoff, H. J. & Klare, J. P. In vivo EPR on spin labeled colicin A reveals an oligomeric assembly of the pore-forming domain in E. coli membranes. *Phys Chem Chem Phys* **17**, 4875-4878 (2015). https://doi.org:10.1039/c4cp05638h

159 Bhakdi, S. *et al.* A guide to the use of pore-forming toxins for controlled permeabilization of cell membranes. *Med Microbiol Immunol* **182**, 167-175 (1993). https://doi.org:10.1007/BF00219946

160 Walev, I. *et al.* Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O. *Proc Natl Acad Sci U S A* **98**, 3185-3190 (2001). https://doi.org:10.1073/pnas.051429498

161 Kotnik, T., Rems, L., Tarek, M. & Miklavcic, D. Membrane Electroporation and Electropermeabilization: Mechanisms and Models. *Annu Rev Biophys* **48**, 63-91 (2019). https://doi.org:10.1146/annurev-biophys-052118-115451

162 Theillet, F. X. *et al.* Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature* **530**, 45-50 (2016). https://doi.org:10.1038/nature16531

163 Limatola, A. *et al.* Time-Resolved NMR Analysis of Proteolytic alpha-Synuclein Processing in vitro and in cellulo. *Proteomics* **18**, e1800056 (2018). https://doi.org:10.1002/pmic.201800056

164 Serber, Z. *et al.* Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc* **1**, 2701-2709 (2006). https://doi.org:10.1038/nprot.2006.181

165 Martorana, A. *et al.* Probing protein conformation in cells by EPR distance measurements using Gd3+ spin labeling. *J Am Chem Soc* **136**, 13458-13465 (2014). https://doi.org:10.1021/ja5079392

166 Yang, Y. *et al.* A Reactive, Rigid Gd(III) Labeling Tag for In-Cell EPR Distance Measurements in Proteins. *Angew Chem Int Ed Engl* **56**, 2914-2918 (2017). https://doi.org:10.1002/anie.201611051

167 Cohen, S. N., Chang, A. C. & Hsu, L. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. *Proc Natl Acad Sci U S A* **69**, 2110-2114 (1972). https://doi.org:10.1073/pnas.69.8.2110

168 Bleicken, S. *et al.* gem-Diethyl Pyrroline Nitroxide Spin Labels: Synthesis, EPR Characterization, Rotamer Libraries and Biocompatibility. *ChemistryOpen* **8**, 1035 (2019). https://doi.org:10.1002/open.201900232

169 Karthikeyan, G. *et al.* A Bioresistant Nitroxide Spin Label for In-Cell EPR Spectroscopy: In Vitro and In Oocytes Protein Structural Dynamics Studies. *Angew Chem Int Ed Engl* **57**, 1366-1370 (2018). https://doi.org:10.1002/anie.201710184

170 Azarkh, M., Okle, O., Eyring, P., Dietrich, D. R. & Drescher, M. Evaluation of spin labels for in-cell EPR by analysis of nitroxide reduction in cell extract of Xenopus laevis oocytes. *J Magn Reson* **212**, 450-454 (2011). https://doi.org:10.1016/j.jmr.2011.07.014

171 Wang, X. W. *et al.* Stabilizing Nitroxide Spin Labels for Structural and Conformational Studies of Biomolecules by Maleimide Treatment. *Chemistry* **29**, e202301350 (2023). https://doi.org:10.1002/chem.202301350

172 Yang, Y. *et al.* High Sensitivity In-Cell EPR Distance Measurements on Proteins using an Optimized Gd(III) Spin Label. *J Phys Chem Lett* **9**, 6119-6123 (2018). https://doi.org:10.1021/acs.jpclett.8b02663

173 Ward, R. *et al.* EPR distance measurements in deuterated proteins. *J Magn Reson* **207**, 164-167 (2010). https://doi.org:10.1016/j.jmr.2010.08.002

174 El Mkami, H. & Norman, D. G. EPR Distance Measurements in Deuterated Proteins. *Methods Enzymol* **564**, 125-152 (2015). https://doi.org:10.1016/bs.mie.2015.05.027