Using the noncanonical metallo-amino Acid [Cu(II)(2,2'-Bipyridin-5-yl)]-alanine to study the structures of proteins

Leandro C. Tabares, Davis T. Daniel, José Luis Vázquez-Ibar, Cyrille Kouklovsky, Valérie Alezra, Sun Un

Using the Noncanonical Metallo-Amino Acid [Cu(II)(2,2'-bipyridin-5-yl)]-alanine to Study the Structures of Proteins

Leandro C. Tabares[†], Davis T. Daniel[†], José Luis Vázquez-Ibar[†], Cyrille Kouklovsky[§], Valérie Alezra[§] and Sun Un[†]*

[†]Institute for Integrative Biology of the Cell, Department of Biochemistry, Biophysics and Structural Biology, Université Paris-Saclay, CEA, CNRS UMR 9198, CEA-Saclay, Gif-sur-Yvette F-91198, France.

[§]Institut de Chimie Moléculaire et des Matériaux d'Orsay (ICMMO), Université Paris-Saclay, CNRS, Orsay F-91405, Cedex, France.

AUTHOR INFORMATION

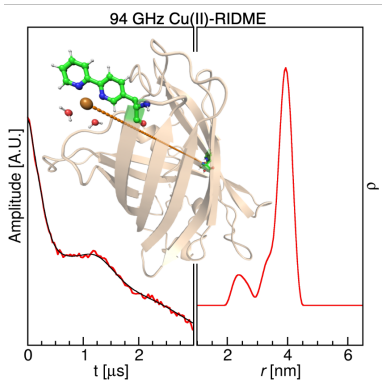
Corresponding Author

* leandro.tabares@cea.fr

ABSTRACT

Genetic code expansion allows modification of the physical and chemical properties of proteins by the site-directed insertion of noncanonical amino-acids. Here we exploit this technology for measuring nanometer scale distance in proteins. (2,2'-Bipyridin-5-yl)alanine was incorporated into the green fluorescent protein and used as anchoring point for Cu(II) to create a spin-label. The incorporation of (2,2'-bipyridin-5-yl)alanine directly into the protein resulted in high affinity binding site for Cu(II) capable of outcompeting other binding positions in the protein. The resulting Cu(II)-spin label is very compact and not larger than a conventional amino acid. By using 94 GHz EPR Pulse Dipolar Spectroscopy we have been able to determine accurately the distance between two of such spin-labels. Our measurements revealed that GFP dimers can adopt different quaternary conformations. The combination of spin-labeling using paramagnetic nonconventional amino-acid with high-frequency EPR techniques resulted in a sensitive method for studying the structures of proteins.

TOC GRAPHICS



KEYWORDS: RIDME, High-field EPR, genetic code expansion, Cu(II) spin-label, structural biology.

The technology for genetically encoding noncanonical amino-acids (ncAA) using exogenous engineered tRNA/aminoacyl-tRNA synthetase (tRNA/aaRS) pairs can site-specifically endow a protein with a new physical or chemical property in ways that conventional site-directed mutagenesis using the 22 canonical amino-acids cannot.[1] This genetic code expansion technology has the potential for greatly diversifying methods for paramagnetic spin-labeling of proteins for magnetic resonance applications. Such spin-labels can then be coupled with Electron Paramagnetic Resonance (EPR) techniques to study the tertiary and quaternary structures of proteins[2,3] by exploiting the magnetic dipolar interaction between two such labels. Two commonly used techniques are DEER[4–6] (Double Electron-Electron Resonance) and RIDME[7,8] (Relaxation Induced Dipolar Modulation Enhancement). Conventional site-directed spin labeling (SDSL) site-specifically introduces cysteine residues that are subsequently linked to nitroxide radicals using sulfur chemistry. This works best when the cysteines are unique and do not physically or functionally affect the protein. Using genetic-code expansion technology, the ncAA azido-phenylalanine has been used, instead of cysteine, to attach a spin-label using azido-alkyne click-chemistry.[9,10] Since it is bio-orthogonal, azido-phenylalanine provides a biologically and chemically unique labeling site. A conceptually simpler and more elegant approach has been to express the targeted protein directly with a paramagnetic ncAA. The first demonstration of this approach was the genetic incorporation of the nitroxide radical 2,2,5,5-tetramethyl-pyrrolin-1-oxyl amino acid into *E. coli* thioredoxin.[11] More recently, 3-pyrazolyltyrosine (PyTyr) has been introduced into the cyan fluorescent protein and the histidine kinase response-regulator CheY. Cu(II) chelation to the ncAA resulted in the PyTyr:Cu(II) spin-label.[12] DEER was used to measure the distance between a PyTyr:Cu(II) center and a nitroxide label attached to a cysteine residue.[12]

An alternative approach has been to use classical site direct mutagenesis to encode the HXXXH (dHis) motif optimally placed in helical structures that can form a high affinity rigid Cu(II) binding site.[13–15] dHis:Cu(II) spin labels have been demonstrated to yield RIDME traces with narrow distance distributions and good sensitivity.

The genetic incorporation of (2,2'-bipyridin-5-yl)alanine (BpyA) was the first demonstration of the use of genetic-code expansion to introduce a metal-binding ncAA into a protein.[16] Bipyridine forms stable metal complexes and binds Cu(II) more strongly than any of the canonical amino-acids.[17,18] BpyA can be viewed as a compact bio-orthogonal single amino-acid version of the dHis motif. Using the pEVOL-BipyA plasmid developed by Schultz and coworkers,[16,19] we have for the first time encoded BypAs into proteins as spectroscopic labels. In particular, BypAs were introduced into the monomeric ultra-stable variant of the green fluorescent protein (muGFP)[20]. Cu(II) were bound to form the BypA:Cu(II) spin-labels. Although GFPs are typically dimeric, muGFP was designed to be monomeric up to concentrations of 10 mg mL^{-1} or about 360 μM.[20] By leveraging the sensitivity benefits of high microwave frequencies, we have been able to obtain high quality 94 GHz RIDME traces from BypA:Cu(II) labeled muGFP samples that were only 1 μL in volume having typical protein concentrations of less than 200 μM. Equally important, these measurements shed light on the existence and structure of muGFP dimers even at low concentrations. We demonstrate that the use of BypA:Cu(II) and high-frequency RIDME is an effective and sensitive approach for examining tertiary and quaternary structures of proteins.

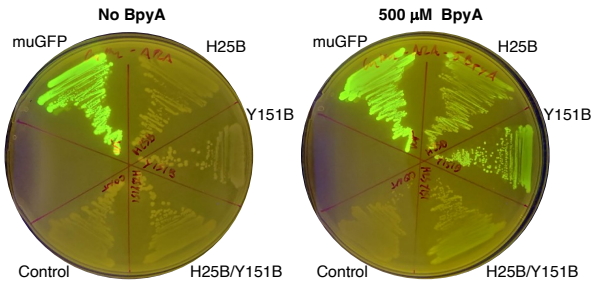


Figure 1. Fluorescence image of *E. coli* cells co-transformed with pEVOL-BipyA and the expression plasmids for muGFP and the H25B, Y151B and H25B/Y151B variants. The empty pET24a plasmid was used as negative control. The cells were grown on LB-agar media complemented with 0.1 mM IPTG, 0.025% arabinose and 0 (left) and 500 mM (right) BpyA.

Expression plasmids for H25B, Y151B and H25B/Y151B (where B is denotes BpyA) mutants of muGFP were prepared and used to co-transform *E. coli* BL21-DE3 cells together with the pEVOL-BipyA plasmid to singly and doubly incorporate BpyA. As expected, the green fluorescence of the cells carrying the mutant plasmids was only observed in media supplemented with BpyA amino acid (Figure 1). The detection of the fluorescence was not only indicative of the BpyA incorporation but also of the correct folding of the muGFP which is required for the chromophore formation. The fluorescence of the H25B and Y151B variants were lower than that of the muGFP. H25B/Y151B had the lowest. Purification studies showed that this was a consequence of the protein expression level rather than misfolding. The yields of Y151B, H25B and H25B/Y151B were 50%, 25% and 10% with respect to the unmodified muGFP. The dependence of the expression level with the position[21,22] and the number of incorporated ncAA is a common occurrence that is not well understood[23]. To best of our knowledge, this is the first time two BpyA have been successfully incorporated into a single protein. Even with diminished

expression levels sufficient amounts of the ncAA labeled proteins were obtainable for spectroscopic applications.

The UV-visible spectra of the purified protein confirmed the BpyA incorporation (Figure 2). The calculated extinction coefficient for muGFP at 280 nm is 20.4 $mM^{-1}cm^{-1}$ and the measured for 2,2'-bipyridine[24], 13 $mM^{-1}cm^{-1}$. The ratios of the absorbance at 280 and 488 nm were: 0.41 for muGFP, 0.68 for H25B and Y151B, and 0.95 for H25B/Y151B, consistent with the incorporation of one BpyA residue in the single mutants and two in the double mutant. An important advantage of using 2,2'-bipyridine is that its complexation to copper, as well as other divalent metals such Co(II) and Ni(II), can be simply and accurately quantified by following the changes in UV-visible absorption at 280, 305 and 317 nm (Figure 2).[25,26] By contrast, binding to the dHis and PyTyr are more difficult to follow. Quantification of Cu(II) loading is important given its propensity to promiscuously binding to proteins. Titrations of the BypA labeled variants with Cu(II) were conducted following the change in absorption at 317 nm (Figure 2, inset). To reach a plateau, 0.8 equivalents of Cu(II) per monomer was required for the H25B variant and 0.9 for the Y151B and 1.7 for the H25B/Y151B. The lower occupancy of H25B was likely due to the presence of Fe(II) that already occupied some of the BpyA centers. This was evident from the significant absorbance at 305 and 317 nm observed in the freshly purified H25B (Figure 2, black trace). Fe(II) occupancy was estimated to be 20%. However, the occupancy depended on the protein expression conditions (see Supporting Information for details). Spurious Fe(II) binding to BpyA has been previously observed.[27,28]

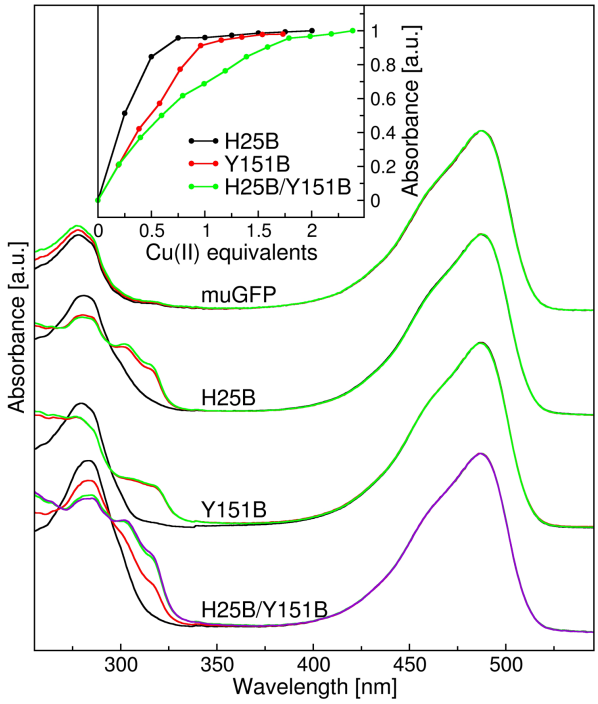


Figure 2. UV-visible absorption spectra of muGFP and BpyA labelled variants. The spectra were taken in the presence of 0 (black), 1 (red), 2 (green) and 3 (violet) equivalents of Cu(II) per protein monomer. Spectra have been normalized to the intensity of the 488 nm GFP chromophore absorption. The inset shows the variation in absorbance at 317 nm as a function of the added equivalents of Cu(II) per protein monomer at pH 6 normalized to unity.

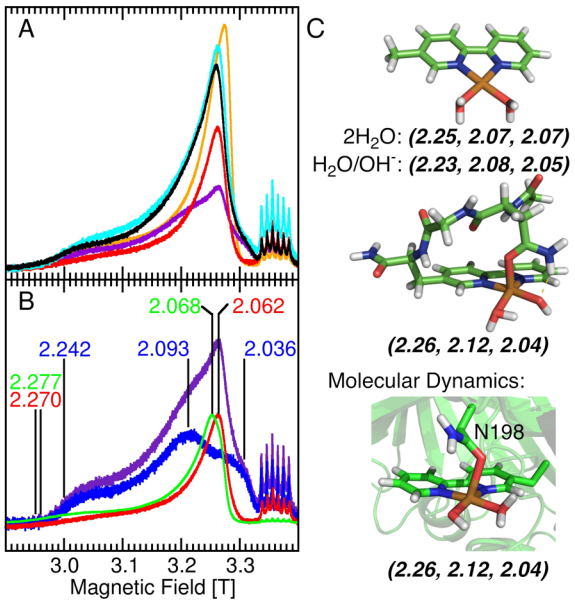


Figure 3. (A) The 30 K 94 GHz field-swept spin echo detected EPR spectra of 50 μM H25B:Cu(II)/Y151B:Cu(II) (black); 50 μM H25B:Cu(II) (red); 50 μM Y151B:Cu(II) (purple) and 200 μM muGFP:Cu(II) (orange). The cyan trace is the sum of the Y151B:Cu(II) and H25B:Cu(II) spectra. The six line Mn(II) spectrum centered at 3.36 T arises from the microwave cavity and provided a convenient magnet-field and amplitude standard. (B) The difference (blue) of the Y151B:Cu(II) spectrum (purple) and a scaled +10 G shifted H25B:Cu(II) (red) spectrum that revealed a second component of the Y151B:Cu(II) spectrum that was 0.7 of the integrated intensity. A spectrum of 500 μM 2,2'-bipyridine:Cu(II) (green) is also shown for comparison. The principal *g*-values of the BypA:Cu(II) centers are shown in the same colors as their spectra. The g_x values were obtained from the 9 GHz spectra shown in the Figure S1. (C) The models of the muGFP BypA:Cu(II) centers and their DFT calculated Cu(II) *g*-tensors: the simple (Cu[(2,2'-bipyridine)$(H_2O)_2$])^2 (top), NGB:Cu(II) model based on the protein sequence (middle) and a model based on molecular dynamics calculations (bottom) in which asparagine-198 donates an axial ligand.

Figure 3 shows 94 GHz field-swept echo-detected EPR spectra of the H25B:Cu(II), Y151B:Cu(II) and H25B:Cu(II)/Y151B:Cu(II) proteins. The three turning points in each spectra

correspond to the three Cu(II) principal *g*-values ($g_{x,y,z} = h\nu/(\gamma B_{x,y,z})$ and $g_x \geq g_y \geq g_z$). The hyperfine structure seen in the conventional 9 GHz EPR spectra (Figure S1) was not resolved. The effects of *g*-straining or distribution in *g*-values appeared to be particularly strong in the g_x region. The spectrum of H25B:Cu(II)/Y151B:Cu(II) was reproduced by the sum of those of H25B:Cu(II) and Y151B:Cu(II) (Figure 3A and S1A). The H25B:Cu(II) spectrum was axial, with principal *g*-values of 2.270, 2.062 and 2.062, indicative of a square-planar Cu(II) centers.[29,30] Y151B:Cu(II) could be decomposed into an axial H25B:Cu(II)-like (30% of the total intensity) component and another (70%) that was anisotropic with principal *g*-values of 2.242, 2.091 and 2.036. The Y151B:Cu(II) shown in Figures 3A and B were representative. The actual compositions varied from sample to sample. The g_z value of 2.24 with an effective $^{63/65}$Cu hyperfine coupling ($|A_{eff}|$) of 440 MHz (1470×10^{-4} cm^{-1}) (Figure 3B and S1B), compared to 540 MHz for H25B:Cu(II)) appeared to be an unusual combination, not corresponding to any of the previously values measured for Cu(II) complexes of bipyridine.[31] This suggested that the metal ligand sphere of the BypA:Cu(II) label was modified by its interactions with the protein. The sample-to-sample variation in the two components indicated that these interactions were likely to be weak.

We calculated the principal *g*-values of $(Cu[(2,2'\text{-bipyridine})(H_2O)_2])^{2+}$ (see Supporting Information). These qualitatively agreed with the measured values for 2,2'-bipyridine:Cu(II) and H25B:Cu(II) (Figure 3C). The *g*-values of a Cu(II) ion bound to the $N_{149}V_{150}B_{151}$ region as shown in Figure 3C exhibited an anisotropy similar to that of the anisotropic Y151B:Cu(II) component. Molecular dynamics (MD) calculations indicated that asparagine-198 could also bind to the Cu(II) ion. The calculated g-values of this arrangement were only negligibly different (see Supporting Information). These calculations provided support that Y151B:Cu(II) had a square pyramid

structure involving protein interaction. Further studies are in progress to ascertain experimental evidence.

The unlabeled muGFP also possessed a copper-binding site. In the presence of 1 equivalent of Cu(II), it had an axial EPR spectrum (orange trace, Figure 3A) that was distinct from those of the BypA:Cu(II). ^{14}N ELDOR-NMR spectroscopy showed that this binding site had at least one histidine (see Figure S3 and Supporting Information). In brief, we observed a distinct 2 MHz split doublet in ^{14}N ELDOR-NMR spectrum of the this native Cu(II) site that is consistent with the hyperfine interaction of Cu(II) with the distal ring nitrogen of a histidine ligand that is not observed for the BypA:Cu(II) centers. This along with the optical data indicated that the BypA sites outcompeted this native site for Cu(II) and the latter did not interfere with the BypA:Cu(II)-based measurements.

Figure 4A shows the 94 GHz Cu(II) RIDME traces from 1 μL samples of single and doubly BypA:Cu(II) labeled muGFP variants as well as muGFP:Cu(II) itself. The time trace of the 50 μM H25B:Cu(II)/Y151B:Cu(II) doubly labeled variant exhibit strong dipolar modulation with a 50% modulation-depth and was obtained in less than 4 h. Even after a 10-fold dilution, analyzable dipolar modulations were readily detectable (< 8 h). These RIDME dipolar modulations build-up during the mixing time interval in the pulse sequence during which the detected spins experience a change in their local magnetic field caused by the longitudinal relaxation of the spins to which they are coupled. For the simple case where the relaxation is exponential with time constant T_1, the modulation-depth is given by

$$\Delta = \frac{Q_{exp}}{2}\left(1 - \exp\left(\frac{-t_{mix}}{T_1}\right)\right) \qquad \text{Eq. 1}$$

where Q_{exp} has been include to account for the possibility that not all detected Cu(II) experience a dipolar interaction with another Cu(II).[32] For the H25B:Cu(II)/Y151B:Cu(II) variant, Q_{exp} is unity.

The T_1 were essentially the same for all cases (Figure S2). The RIDME traces shown in Figure 4 were obtained with a 32 μs mixing time or approximately1.5 T_1. This mixing time was a reasonable balance between good modulation depths, which increases with the mixing time, and significant loss in signal-to-noise which invariably occurs with increasing mixing time.[33] The 30 K temperature was a compromise between short T_1 allowing fast shot-repetition times (153 μs) for efficient signal averaging and sufficiently long phase memory times to insure the 3 μs RIDME measurement window. Evidently, these compromises did not significantly affect the quality of the measurement.

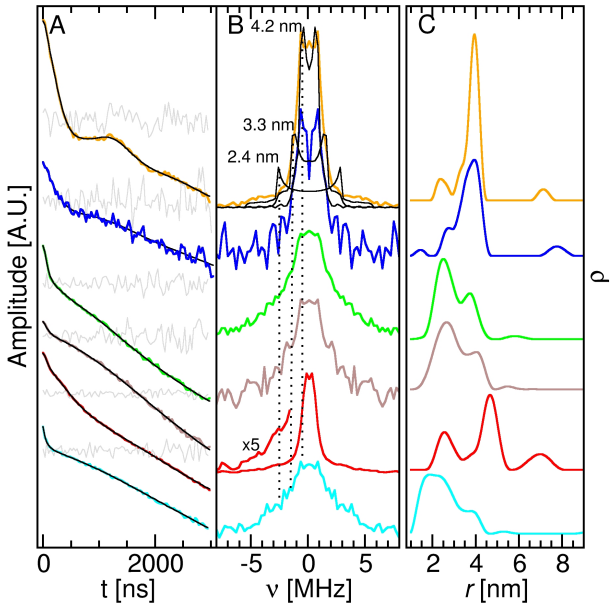


Figure 4. (A) 94 GHz Cu(II) RIDME traces of 50 (orange) and 5 μM (blue) H25B:Cu(II)/Y151B:Cu(II); 250 (green) and 50 μM (brown) Y151B:Cu(II); 200μM H25B:Cu(II) (red); and 200 μM Cu(II) added to 200 μM muGFP (cyan). The traces were obtained at the maximum of the corresponding field-swept echo-detected spectrum. (B) The Fourier transforms of each RIDME trace after subtraction of stretch exponential fits of their backgrounds. The black traces are the calculated dipolar spectrum of two ½-spin separated by the distances indicated. (C) The modified-Tikhonov distance distribution of the RIDME traces. Fits (thin black lines) using the same background fits are superimposed in panel A and the horizontal gray traces are the

differences between the modified-Tikhonov fits and measurements scaled by a factor of 5, except for 5 μM H25B:Cu(II)/Y151B:Cu(II) which has been scaled by 2.5.

Intrinsic to these RIDME measurements, where only a small fraction of the Cu(II) spins are excited and detected, is the possibility of orientation selection. This is because each position in the Cu(II) EPR spectra arises from a unique set of orientations of each BpyA:Cu(II) with respect to the magnetic field. In such cases, the orientation dependence of the dipolar interaction may not be sufficiently sampled to reliably extract the distance between the Cu(II) centers. The use of high magnetic-fields potentially exacerbates this problem since a smaller set of Cu(II) are excited and detected. However as further discussed in the Supporting Information, this did not seem to be a significant problem presumably due to the rotational flexibility afforded by the two C_β bonds that link the BypA:Cu(II) label to protein backbone. MD calculation based on the muGFP crystal structure were used to ascertain the conformational space spanned by a BpyA:Cu(II) label. In these MD simulations, the label was modeled as $[Cu(BpyA)(H_2O)_2]^{2+}$ and the less bulky $[Cu(BpyA)]^{2+}$ (see Supporting Information for details). As expected, the simulations showed that the $[Cu(BpyA)]^{2+}$ was less sterically constrained, with $\vec{r}_{C_\alpha - Cu(II)}$ (the vector connecting the Cu(II) to C_α of the same residue) spanning a wider angular space than the hydrated complex (Figure 5A). Both forms of the BpyA:Cu(II) spin-label were spatially very compact. The conformational space spanned by the Cu(II) atom was similar in size to that of a phenoxyl oxygen atom of a tyrosine amino-acid (Figure 5A). In spite of this compactness, the interatomic $\vec{r}_{N-Cu(II)}$ vector subtended a large portion of an octant of a sphere (Figure S5). It is likely that due to the distributed nature of the Cu(II) Zeeman interaction this coverage was even greater than predicted by the MD simulations. This meant that the RIDME measurements likely sampled a large range of the

orientation dependence of dipolar interactions resulting in the lack of any significant orientation selection effects in the measurements (see Supporting Information).

The dipolar spectra of the BypA:Cu(II) labeled variants were obtained by Fourier transforming (FT) their RIDME traces (Figure 4B). Superimposed on the FT of the H25B:Cu(II)/Y151B:Cu(II) variant RIDME time trace are three calculated dipolar spectra (Pake patterns) corresponding to Cu(II)-Cu(II) distances of 4.2, 3.0, and 2.4 nm. The dipolar spectra of all three BypA:Cu(II) labeled variants appeared to have similar components. The FT do not allow for complete quantitative analyses of the dipolar interactions. The most common approach is based on Tikhonov regularization. Rather than the conventional approach where the background is fitted and then subtracted or, more properly, divided from the measured traces, we implemented and use the approach of Fábregas Ibáñez and Jeschke[34] which incorporates the background directly into the Tikhonov regularization, specifically the modified $K_{\sqrt{B}}$ kernel, the details for which are found in their publication and which hereon will be referred to as the modified-Tikhonov calculation. A single stretch exponential model of the background of the RIDME time traces (Table S2) yielded the better fits and more stable solutions than polynominals.[35,36] The modified-Tikhonov fit of these traces are shown in Figure 4 and conventional Tikhonov regularization in Figure S6. For simplicity, a brute-force grid search that scanned the zero time-point, the region used to fit the stretch background and regularization parameter was used. For each zero time-point and fitting region, a stretch exponential was fit to the data which along with the chosen regularization parameter were then used for the modified-Tikhonov calculation. Shown in Figures 4C are the fits with the smallest RMS deviation from the measured data. The regularization parameters were typically 10 and corresponded to the corner of the l-curve. The numerical "noise" can be significant when the background is simply divided adversely affecting the analysis. This was much less

evident with the modified-Tikhonov approach. Otherwise, as seen on Figure 4A the fits were free of systematic deviations.

The distance distributions returned by the modified-Tikhonov regularization (Figure 4C) were consistent with what we had qualitatively surmised from the Fourier transforms. The doubly and singly labeled variants had Cu-to-Cu distances corresponding approximately 2.5, 3.5 and 4 nm. In nearly all cases, there were low probability distances above 6 nm. However, these were not stable with respect to the scanned parameters. The presence of this feature likely arose from the inadequacies of modeling the background as a single stretch exponential.[37] Moreover, the 3 μs dipolar evolution time was insufficiently long to make this feature reliable. The H25B:Cu(II)/Y151B:Cu(II) 3.9 nm distance was the expected distance based on the separation of H25 and Y151 in the muGFP crystal structure (Figure 5B). Since the hyperfine and optical measurements precluded the presence of Cu(II) bond to sites other than a BpyA, the simplest explanation for the 2.5 and 3.5 nm H25B:Cu(II)/Y151B:Cu(II) features, as well as the RIDME modulations of the singly label variants, was the presence of significant concentrations of dimers. The Y151B:Cu(II) variant exhibited a drop in modulation depth from 20% to 8% with a decrease in protein concentration (Figure 4A), consistent with a shift in the monomer-dimer equilibrium favoring monomers. Nonetheless, dimers were present even at 50 μM, in contrast to previous observations that muGFP is predominately monomeric below 360 μM [20]. One would have also expected the background of the more concentrated sample to decay significantly faster since it results from a greater number of non-specific dipolar interactions between Cu(II) ions.[36] However, only a small increase was observed (Table S2). This may have in part been due to compensating changes in the mono-dimer equilibrium. Further study is required. The 2.6 and 3.8 nm distances found for the Y151B variant suggested the presence of two different quaternary structures.

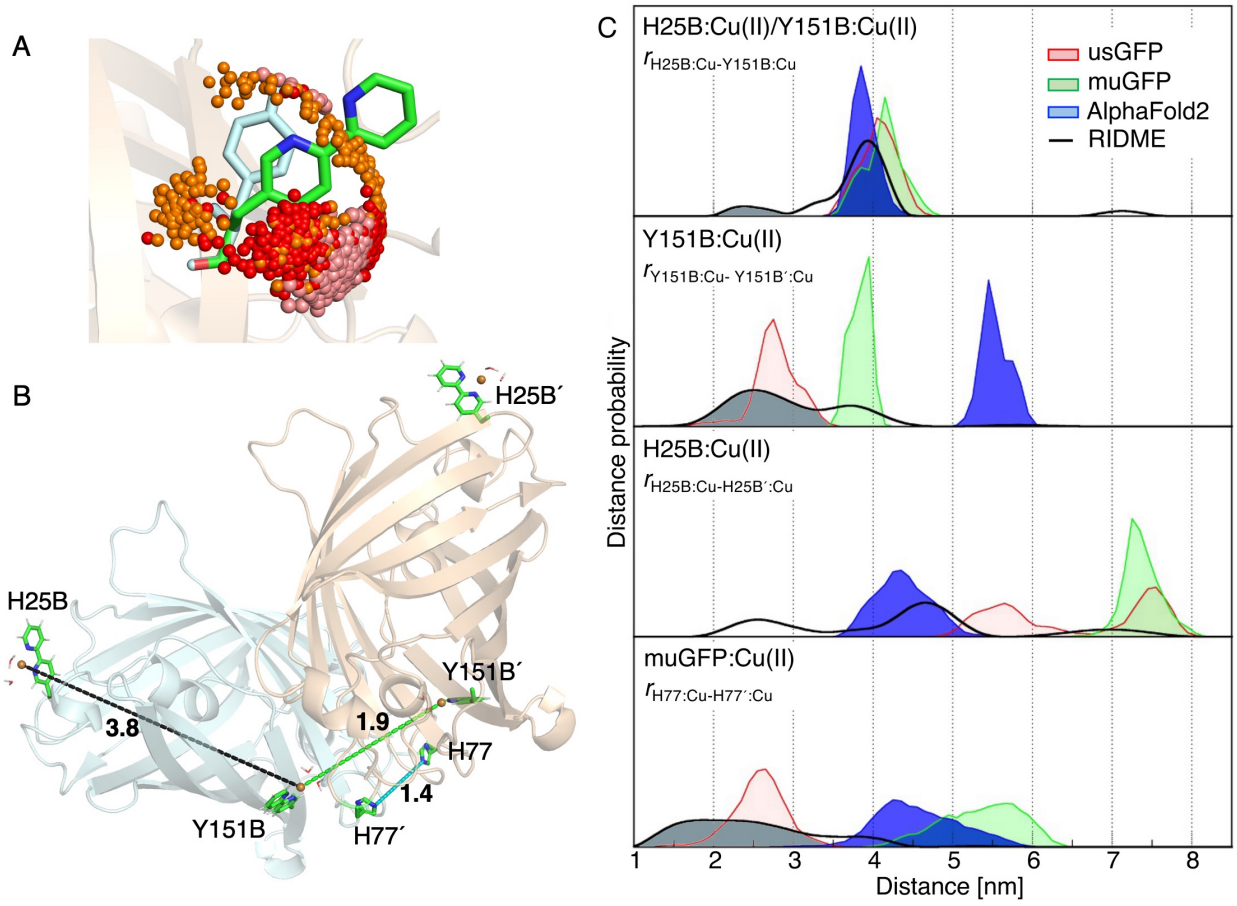


Figure 5. MD simulations: (A) The positions spanned by the phenoxyl oxygen (pink spheres) of Y151 (outlined in light blue) and the Cu(II) ion of $[Cu(II)BpyA(H_2O)_2]^{2+}$ (red spheres) and $[Cu(II)BpyA]^{2+}$ (orange spheres) models of the sidechain of Y151B:Cu(II) (outlined in green and blue). (B) Structural model of H25B/Y151B with the relevant distances indicated based on the dimeric usGFP structure. (C) The MD distance profiles for monomeric $r_{H25B:Cu-Y151B:Cu}$ and dimeric $r_{Y151B:Cu-Y151B´:Cu}$, $r_{H25B:Cu-H25B´:Cu}$ and $r_{H77:Cu-H77´:Cu}$ based on the usGFP (red shaded) and muGFP (green shaded) crystal structures and an AlphaFold2 structure (blue shaded). The black traces show the corresponding RIDME results. All distance profiles have been normalized to their total probabilities. The grey shaded regions indicate distances arising consistent with the usGFP quaternary structure.

MD simulations were used to develop models of the spin-labeled muGFP dimers. Three different quaternary arrangements of the muGFP variants were considered: the crystallographic form of muGFP (PDB code: 5JZL) in which the β-barrels of the two monomers present in the asymmetric unit are anti-parallel; the crystallographic structure of the dimeric parent usGFP (PDB code: 5JZK) in which the β-barrels are nearly perpendicular; and an AlphaFold2[38–40] model in which they are parallel (Figure S7 and Supporting Information). In these structures there was a histidine rich region centered around H77 (H77, H81, H199, and H231' of the other neighboring monomer in the usGFP model) that was an attractive candidate for the Cu(II) binding site in the unlabeled muGFP. Figure 5C shows the MD distance distribution for the dimeric $r_{\mathrm{H77:Cu\text{-}H77':Cu}}$ for the three different models. Unlike the others, the usGFP-based distance profile had approximately the same center as the measured by RIDME. The narrower MD distribution likely reflected the omission of the different binding configurations that involve the other histidines present in this region. Their inclusion would presumably increase the breadth of the distance distribution.

The MD profile for $r_{\mathrm{H25B:Cu(II)\text{-}Y151B:Cu(II)}}$ for the H25B:Cu(II)/Y151B:Cu(II) variant matched the prominent 3.9 nm feature in measured distance distribution independently of the quaternary model (Figure 5C). The smaller 2.4 nm feature of this variant had similar counterpart in the distance distribution of the singly labeled Y151B:Cu(II) variant (grey shaded region in Figure 5C). These coincided with the MD profile of $r_{\mathrm{Y151B:Cu(II)\text{-}Y151B':Cu(II)}}$ based on the usGFP. The 1 to 3.5 nm portion of the H77-H77′ distance profile of the unlabeled muGFP is also consistent with the usGFP quaternary structure. In contrast the 3.7 nm $r_{\mathrm{Y151B:Cu(II)\text{-}Y151B':Cu(II)}}$ profile matched the antiparallel structure of muGFP. The MD $r_{\mathrm{H25B:Cu\text{-}H25B'}}$ profile for the of H25B:Cu(II) variant based on usGFP and muGFP structures did not agree with the experimental data. The AlphaFold2 model-based distance profile could account for the longer half of the experimentally observed bimodal

distribution that had maxima at 2.5 and 4.7 nm (Figure 5C). This suggested the possibility that in this variant some of the dimers had monomers in a parallel configuration.

The comparison of RIDME measurements and MD simulations indicated that different quaternary conformations are taken by the muGFP variants. This is not surprising since a single point-mutation can alter the GFP monomer/dimer equilibrium and modify the dimeric structure.[41,42] Given this sensitivity, we cannot exclude the possibility that the BypA mutation or the Cu(II) influences the monomer/dimer equilibrium. However, we do know that for the BypA variants, no BypA:Cu(II):BypA centers that bridge the muGFP monomers were present.[43] The titration experiments showed a nearly stochiometric Cu(II) to BpyA ratio excluding the presence of significant quantities of $[BpyA_2Cu]^{2+}$ and $[BpyA_3Cu]^{2+}$ complexes and any Cu(II) bond to other places in the protein. The crystal structures and MD calculations did not reveal any other sufficiently close candidate ligands, in particular glutamates and aspartates, that could form $[Cu(BpyA)X]^{1+}$ and bridged the two monomers. In the unlabeled variant, one of the histidines around H77 does come from the neighboring monomer in the usGFP model; hence, Cu(II) may have promoted dimer formation in this case.[44]

The binding of Cu(II) to unintended sites in the protein is a common issue of Cu(II)-based spin labels.[12,13] Based on our results, it is likely that BpyA is capable of outcompeting most forms of unintended Cu(II) binding site(s) such as those formed by multiple histidines. Moreover, overloading of the BypA sites can be quantitatively avoided using simple optical spectroscopy. The genetic incorporation of metal-chelating amino acids appears to provide excellent opportunities for novel spin-labels with ideal properties. Here we demonstrated this using BpyA:Cu(II), but other metals, notably Co(II), will also work as spin-labels. BpyA:Cu(II) is chemical stable and structurally compact. MD simulation showed that it does not occupy much

larger space than a tyrosine. Yet, it is sufficiently rotationally flexible that orientation selection does not appear to affect distance measurement. The compactness of the BpyA:Cu(II) label also insures that its contribution to the uncertainty in the distance measurements is minimal. The use of high microwave frequencies to perform Pulse Dipolar Spectroscopy measurements on these centers is not an obvious choice since the width of the Cu(II) EPR spectrum range increase linearly with frequency. In spite of this, we have shown that using 94 GHz RIDME it is possible to measure with good accuracy the dipolar interaction between two labels even for protein concentrations as low as 5 μM using only a 1 μL of sample. Here we demonstrated that the combination of the paramagnetic non-conventional amino acid BpyA:Cu(II) with high-frequency, high-magnetic-field, EPR, resulted in a sensitive methodology for the study of the structure of proteins.

ASSOCIATED CONTENT

Supporting Information Available: 9 GHz cw-EPR spectra of BpyA labeled variants of muGFP, characterization of the native Cu(II) binding site in muGFP, DFT calculations on a potential muGFP:Cu(II) site, orientation selection, T_1 measurements, background fits parameters of the RIDME traces, conventional Tikhonov regularization using DeerAnalysis, modeling of the quaternary structures muGFP and experimental and computational methods.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

Davis T. Daniel current address: Institute for Energy and Climate Research (IEK-9), Forschungzentrum Jülich GmbH, Jülich, 52424, Germany

Corresponding Author: Leandro C. Tabares, Bat 532 Pce 215, F-91191 Gif-sur-Yvette, France, phone: +33-169087579, email: leandro.tabares@cea.fr.

ACKNOWLEDGMENT

We are grateful to Dr. Peter Schultz (Scripps Research) for his generous gift of the pEVOL-BipyA plasmid. We acknowledge Tania Tibiletti (I2BC) for her assistance with 9 GHz measurements. This work was financially supported by the ANR InVivoNanoSpin (ANR-20-CE11-0002) and the Ile-de-France DIM-ELICIT program. We benefited from the biophysics platform of the Institute for Integrative Biology of the Cell supported by French Infrastructure for Integrated Structural Biology (FRISBI, ANR-10-INBS-0005) and the Infrastructure for Biology, Health and Agronomy (IBISA).

REFERENCES

(1) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu Rev Biochem* **2010**, *79*, 413–444.

(2) Jeschke, G. DEER Distance Measurements on Proteins. *Annu Rev Phys Chem* **2012**, *63* (1), 419–446.

(3) Schiemann, O.; Heubach, C. A.; Abdullin, D.; Ackermann, K.; Azarkh, M.; Bagryanskaya, E. G.; Drescher, M.; Endeward, B.; Freed, J. H.; Galazzo, L.; Goldfarb, D.; Hett, T.; Esteban Hofer, L.; Fábregas Ibáñez, L.; Hustedt, E. J.; Kucher, S.; Kuprov, I.; Lovett, J. E.; Meyer, A.; Ruthstein, S.; Saxena, S.; Stoll, S.; Timmel, C. R.; di Valentin, M.; Mchaourab, H. S.; Prisner, T. F.; Bode, B. E.; Bordignon, E.; Bennati, M.; Jeschke, G. Benchmark Test and Guidelines for DEER/PELDOR Experiments on Nitroxide-Labeled Biomolecules. *J Am Chem Soc* **2021**, *143* (43), 17875–17890.

(4) Milov, a. D.; Salikhov, K. M.; Shchirov, M. D. Use of the Double Resonance in Electron Spin Echo Method for the Study of Paramagnetic Center Spatial Distribution in Solids. *Fizika Tverdogo Tela*. 1981, pp 975–982.

(5) Larsen, R. G.; Singel, D. J. Double Electron-Electron Measurement of Electron Disordered Solids Resonance Spin-Echo Spin Pair Separations Modulation : Spectroscopic in Orientationally Disordered Solids. *J Chem Phys* **1993**, *98* (1993), 5134–5146.

(6) Martin, R. E.; Pannier, M.; Diederich, F.; Gramlich, V.; Hubrich, M.; Spiess, H. W. 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* **1998**, *37* (20), 2833–2837.

(7) Kulik, L. V.; Dzuba, S.; Grigoryev, I.; Tsvetkov, Y. D. Electron Dipole-Dipole Interaction in ESEEM of Nitroxide Biradicals. *Chem Phys Lett* **2001**, *343*, 315–324.

(8) 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* **2009**, *201* (1), 48–56.

(9) Widder, P.; Berner, F.; Summerer, D.; Drescher, M. Double Nitroxide Labeling by Copper-Catalyzed Azide-Alkyne Cycloadditions with Noncanonical Amino Acids for Electron Paramagnetic Resonance Spectroscopy. *ACS Chem Biol* **2019**, *14* (5), 839–844.

(10) Abdelkader, E. H.; Feintuch, A.; Yao, X.; Adams, L. A.; Aurelio, L.; Graham, B.; Goldfarb, D.; Otting, G. Protein Conformation by EPR Spectroscopy Using Gadolinium Tags Clicked to Genetically Encoded P-Azido-L-Phenylalanine. *Chem Comm* **2015**, *51* (88), 15898–15901.

(11) Schmidt, M. J.; Borbas, J.; Drescher, M.; Summerer, D. A Genetically Encoded Spin Label for Electron Paramagnetic Resonance Distance Measurements. *J Am Chem Soc* **2014**, *136* (4), 1238–1241.

(12) Merz, G. E.; Borbat, P. P.; Muok, A. R.; Srivastava, M.; Bunck, D. N.; Freed, J. H.; Crane, B. R. Site-Specific Incorporation of a Cu^{2+} Spin Label into Proteins for Measuring Distances by Pulsed Dipolar Electron Spin Resonance Spectroscopy. *J Phys Chem B* **2018**, *122* (41), 9443–9451.

(13) 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* **2015**, *54* (21), 6330–6334.

(14) Ghosh, S.; Lawless, M. J.; Rule, G. S.; Saxena, S. The Cu2+-Nitrilotriacetic Acid Complex Improves Loading of α-Helical Double Histidine Site for Precise Distance Measurements by Pulsed ESR. *J Mag Res* **2018**, *286*, 163–171.

(15) Ackermann, K.; Wort, J. L.; Bode, B. E. Nanomolar Pulse Dipolar EPR Spectroscopy in Proteins: CuII-CuII and Nitroxide-Nitroxide Cases. *J Phys Chem B* **2021**, *125* (20), 5358–5364.

(16) Xie, J.; Liu, W.; Schultz, P. G. A Genetically Encoded Bidentate, Metal-Binding Amino Acid. *Angew Chem Int Ed Engl* **2007**, *46* (48), 9239–9242.

(17) Irving, H.; Mellor, D. H. 1002. The Stability of Metal Complexes of 1,10-Phenanthroline and Its Analogues. Part I. 1,10-Phenanthroline and 2,2′-Bipyridyl. *J Chem Soc (Resumed)* **1962**, 5222–5237.

(18) Williams, D. R. Thermodynamic Considerations in Co-Ordination. Part VII. Solubility of the Histidine–H^+ System and Stability Constants, Free Energies, Enthalpies, and Entropies of Protonation of Histidine and Tryptophan and of Formation of Their Manganese(II), Iron(II), Cobalt(II), Nickel(II), Copper(II), and Zinc(II) Complexes. *J. Chem. Soc. A* **1970**, *0* (0), 1550–1555.

(19) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli. *J Mol Biol* **2010**, *395* (2), 361–374.

(20) Scott, D. J.; Gunn, N. J.; Yong, K. J.; Wimmer, V. C.; Veldhuis, N. A.; Challis, L. M.; Haidar, M.; Petrou, S.; Bathgate, R. A. D.; Griffin, M. D. W. A Novel Ultra-Stable, Monomeric Green Fluorescent Protein for Direct Volumetric Imaging of Whole Organs Using CLARITY. *Sci Rep 2018 8:1* **2018**, *8* (1), 1–15.

(21) Albayrak, C.; Swartz, J. R. Cell-Free Co-Production of an Orthogonal Transfer RNA Activates Efficient Site-Specific Non-Natural Amino Acid Incorporation. *Nucleic Acids Res* **2013**, *41* (11), 5949.

(22) Schinn, S. M.; Bradley, W.; Groesbeck, A.; Wu, J. C.; Broadbent, A.; Bundy, B. C. Rapid in Vitro Screening for the Location-Dependent Effects of Unnatural Amino Acids on Protein Expression and Activity. *Biotechnol Bioeng* **2017**, *114* (10), 2412–2417.

(23) Amiram, M.; Haimovich, A. D.; Fan, C.; Wang, Y. S.; Aerni, H. R.; Ntai, I.; Moonan, D. W.; Ma, N. J.; Rovner, A. J.; Hong, S. H.; Kelleher, N. L.; Goodman, A. L.; Jewett, M. C.; Söll, D.; Rinehart, J.; Isaacs, F. J. Evolution of Translation Machinery in Recoded Bacteria Enables Multi-Site Incorporation of Nonstandard Amino Acids. *Nature Biotechnol* **2015**, *33* (12), 1272–1279.

(24) Nakamoto, K. Ultraviolet Spectra and Structures of 2,2′-Bipyridine and 2,2′,2″-Terpyridine in Aqueous Solution. *J Phys Chem* **1960**, *64* (10), 1420–1425.

(25) Mason, S. F. The Electronic Spectra and Optical Activity of Phenanthroline and Dipyridyl Metal Complexes. *Inorganica Chim Acta Reviews* **1968**, *2* (C), 89–109.

(26) Nakamoto, K. Ultraviolet Spectra and Structures of 2,2′-Bipyridine and 2,2′,2″-Terpyridine in Aqueous Solution. *J Phys Chem* **1960**, *64* (10), 1420–1425.

(27) Mills, J. H.; Khare, S. D.; Bolduc, J. M.; Forouhar, F.; Mulligan, V. K.; Lew, S.; Seetharaman, J.; Tong, L.; Stoddard, B. L.; Baker, D. Computational Design of an Unnatural Amino Acid Dependent Metalloprotein with Atomic Level Accuracy. *J Am Chem Soc* **2013**, *135* (36), 13393–13399.

(28) Bersellini, M.; Roelfes, G. Multidrug Resistance Regulators (MDRs) as Scaffolds for the Design of Artificial Metalloenzymes. *Org Biomol Chem* **2017**, *15* (14), 3069–3073.

(29) Bennett, B.; Kowalski, J. M. EPR Methods for Biological Cu(II): L-Band CW and NARS. *Methods Enzymol* **2015**, *563*, 341.

(30) Solomon, E. I.; Heppner, D. E.; Johnston, E. M.; Ginsbach, J. W.; Cirera, J.; Qayyum, M.; Kieber-Emmons, M. T.; Kjaergaard, C. H.; Hadt, R. G.; Tian, L. Copper Active Sites in Biology. *Chem Rev* **2014**, *114* (7) 3659–3853.

(31) Garribba, E.; Micera, G.; Sanna, D.; Strinna-Erre, L. The Cu(II)-2,2′-Bipyridine System Revisited. *Inorganica Chim Acta* **2000**, *299* (2), 253–261.

(32) Wort, J. L.; Ackermann, K.; Norman, D. G.; Bode, B. E. A General Model to Optimise CuII Labelling Efficiency of Double-Histidine Motifs for Pulse Dipolar EPR Applications. *Phys Chem Chem Phys* **2021**, *23* (6), 3810–3819.

(33) Breitgoff, F. D.; Keller, K.; Qi, M.; Klose, D.; Yulikov, M.; Godt, A.; Jeschke, G. UWB DEER and RIDME Distance Measurements in Cu(II)–Cu(II) Spin Pairs. *J Mag Res* **2019**, *308*, 106560.

(34) Fábregas Ibáñez, L.; Jeschke, G. Optimal Background Treatment in Dipolar Spectroscopy. *Phys Chem Chem Phys* **2020**, *22* (4), 1855–1868.

(35) Abdullin, D.; Brehm, P.; Fleck, N.; Spicher, S.; Grimme, S.; Schiemann, O. Pulsed EPR Dipolar Spectroscopy on Spin Pairs with One Highly Anisotropic Spin Center: The Low-Spin FeIII Case. *Chem Eur J* **2019**, *25* (63), 14388–14398.

(36) Keller, K.; Qi, M.; Gmeiner, C.; Ritsch, I.; Godt, A.; Jeschke, G.; Savitsky, A.; Yulikov, M. Intermolecular Background Decay in RIDME Experiments. *Phys Chem Chem Phys* **2019**, *21* (16), 8228–8245.

(37) Keller, K.; Doll, A.; Qi, M.; Godt, A.; Jeschke, G.; Yulikov, M. Averaging of Nuclear Modulation Artefacts in RIDME Experiments. *J Mag Res* **2016**, *272*, 108–113.

(38) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv* **2022**, 2021.10.04.463034.

(39) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat Methods* **2022**, *19* (6), 679–682.

(40) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

(41) Palm, G. J.; Zdanov, A.; Wlodawer, A.; Gaitanaris, G. A.; Stauber, R.; Pavlakis, G. N. The Structural Basis for Spectral Variations in Green Fluorescent Protein. *Nat. Struct. Biol.* **1997**, *4* (5), 361–365.

(42) Leibly, D. J.; Arbing, M. A.; Pashkov, I.; Devore, N.; Waldo, G. S.; Terwilliger, T. C.; Yeates, T. O. A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **2015**, *23* (9), 1754–1768.

(43) Yang, M.; Song, W. J. Diverse Protein Assembly Driven by Metal and Chelating Amino Acids with Selectivity and Tunability. *Nat Commun* **2019**, *10* (1), 1–11.

(44) Leibly, D. J.; Arbing, M. A.; Pashkov, I.; Devore, N.; Waldo, G. S.; Terwilliger, T. C.; Yeates, T. O. A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **2015**, *23* (9), 1754–1768.