Through-Space Scalar $^{19}F–^{19}F$ Couplings between Fluorinated Non-Canonical Amino Acids for the Detection of Specific Contacts in Proteins

Henry W. Orton,[1‡] Haocheng Qianzhu,[2‡] Elwy H. Abdelkader,[1] Edan I. Habel,[2] Yi Jiun Tan,[1] Rebecca L. Frkic,[1] Colin J. Jackson,[1] Thomas Huber,[2] and Gottfried Otting[1]*

[1] ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

KEYWORDS CF_3 groups, NMR spectroscopy, non-canonical amino acids, proteins, through-space scalar fluorine–fluorine couplings.

ABSTRACT: Fluorine atoms are known to display scalar $^{19}F–^{19}F$ couplings in nuclear magnetic resonance (NMR) spectra when they are sufficiently close in space for non-bonding orbitals to overlap. We show that fluorinated non-canonical amino acids positioned in the hydrophobic core or on the surface of a protein can be linked by scalar through-space $^{19}F–^{19}F$ ($^{TS}J_{FF}$) couplings even if the ^{19}F spins are in the time average separated by more than the van der Waals distance. Using two different aromatic amino acids featuring CF_3 groups, O-trifluoromethyl-tyrosine and 4-trifluoromethyl-phenylalanine, we show that $^{19}F–^{19}F$ TOCSY experiments are sufficiently sensitive to detect TSJFF couplings between 2.5 and 5 Hz in the 19 kDa protein PpiB measured on a 2-channel 400 MHz NMR spectrometer with a regular room temperature probe. A quantitative J evolution experiment enables the measurement of $^{TS}J_{FF}$ coupling constants that are up to five times smaller than the ^{19}F NMR line width. In addition, a new aminoacyl-tRNA synthetase was identified for genetic encoding of N^6-(trifluoroacetyl)-L-lysine (TFA-Lys) and $^{19}F–^{19}F$ TOCSY peaks were observed between two TFA-Lys residues incorporated into the proteins AncCDT-1 and mRFP despite high solvent exposure and flexibility of the TFA-Lys side chains. With the ready availability of systems for site-specific incorporation of fluorinated amino acids into proteins by genetic encoding, $^{19}F–^{19}F$ interactions offer a straightforward way to probe the spatial proximity of selected sites without any assignments of ^{1}H NMR resonances.

INTRODUCTION

The ^{19}F nucleus is attractive for NMR spectroscopy as it has a spin ½ with a high Larmor frequency, and ^{19}F is the sole naturally occurring fluorine isotope. As a drawback, the large chemical shift anisotropy (CSA) associated with ^{19}F causes significant line broadening in macromolecules at high magnetic field strength. Nonetheless, as biological systems generally do not contain fluorine, ^{19}F NMR spectra of proteins are devoid of background resonances and, therefore, ^{19}F NMR has long been shown to be attractive for protein studies following site-selective labeling, either by the incorporation of fluorinated amino acids in lieu of regular amino acids or the use of fluorinated ligand molecules.[1–9] In addition, fluorine atoms occur in about 20% of all commercially available pharmaceutical drugs[10] and about half of the most successful contemporary drug molecules.[11]

An early report of *Lactobacillus casei* dihydrofolate reductase (DHFR) containing 6-fluorotryptophan noted the observation of a through-space scalar $^{19}F–^{19}F$ ($^{TS}J_{FF}$) coupling of 21 Hz between two fluorine atoms, which came about by a fortuitous contact between Trp5 and Trp133 in the three-dimensional (3D) structure of the protein.[12] Indeed, through-space scalar couplings are common between ^{19}F spins that are close in space.[13] They can also be sizeable, as shown by the example of DHFR and, more recently, in a G-quadruplex analogue synthesized with fluorinated arabinoguanosine, which displayed a through-space $^{19}F–^{19}F$ coupling of 43 Hz.[14] While long-range scalar $^{19}F–^{19}F$ couplings in small molecules are commonly observed,[13] the limited size of small molecules allows distinguishing between through-space and conventional through-bond effects often only with the help of DFT calculations.[15–17] In contrast, it is straightforward to position fluorine atoms in a protein such that they are in close spatial proximity yet separated by many covalent bonds. In this situation, any $^{19}F–^{19}F$ couplings can be unambiguously assigned to through-space effects. Notably, through-space scalar couplings always are a manifestation of orbital overlap, so that the distinction between through-space and through-bond couplings is operational in character rather than a matter of fundamental difference.

The observation of through-space couplings is not limited to $^{19}F–^{19}F$ interactions and, in proteins, has also been reported between methyl protons and π bond systems of carbonyl and aromatic carbons,[18] between spins connected by H-bonds,[19–23] or with heavy metals such as cadmium and mercury.[24] In view of the popularity of protein labeling with fluorinated amino acids and the common observation of $^{TS}J_{FF}$ couplings in small molecules, it is surprising that, to the best of our knowledge, through-space $^{19}F–^{19}F$ couplings in proteins have not been reported since the 1978 publication on DHFR.[12] To a large extent, this must be

attributed to the large chemical shift anisotropy associated with ^{19}F spins which, at high magnetic field strength, causes substantial line broadening in high-molecular weight systems such as proteins. Although the CSA relaxation can be alleviated by positioning the fluorine spins in highly flexible chemical groups, such as a CF_3 group at the end of a flexible tether,[5] and this approach has been very successful in monitoring different conformations of integral membrane proteins induced by different ligands,[25–27] it doesn't lend itself to the observation of through-space scalar ^{19}F–^{19}F couplings.

The advent of mutant aminoacyl-tRNA synthetases has made it possible to install a single non-canonical amino acid at a specific site of a protein in response to an amber stop codon.[28,29] With the installation of two fluorinated amino acids, it is possible to detect their proximity by inter-residual ^{19}F–^{19}F NOEs. For example, ^{19}F–^{19}F NOEs were recently demonstrated for two *para*-pentafluorsulfanyl phenylalanine (SF_5-Phe) residues incorporated in the 19 kDa *E. coli* protein *cis-trans* prolyl-isomerase B (PpiB).[30] SF_5 groups, however, are very bulky, which could affect the protein structure, and display large intraresidual ^{19}F–^{19}F couplings in the ^{19}F NMR spectrum, which limit the spectral resolution. This prompted us to investigate the interaction between the CF_3 groups of two *para*-trifluoromethyl-phenylalanine (CF_3-Phe) or *O*-trifluoromethyl-tyrosine (CF_3-Tyr) residues. Aminoacyl-tRNA synthetases for amber suppressor-tRNA have been described previously for both amino acids.[31–33] These aminoacyl-tRNA synthetases are highly efficient and the incorporation of multiple copies of CF_3-Phe residues has been demonstrated.[34]

In the following, we show that the ^{19}F NMR resonances of two CF_3-Phe or CF_3-Tyr residues in PpiB are significantly narrower than those of SF_5-Phe residues, and those of solvent-exposed TFA-Lys residues are even narrower. Remarkably, through-space scalar ^{19}F–^{19}F couplings produced readily observable cross-peaks in ^{19}F–^{19}F DQF-COSY and TOCSY spectra recorded on a 400 MHz NMR spectrometer equipped with a conventional room temperature probe. ^{19}F–^{19}F TOCSY cross-peaks were observed even for couplings much smaller than the ^{19}F NMR line width and their signal to noise ratio was competitive with ^{19}F–^{19}F NOESY cross-peaks. Crystal structures of PpiB mutants containing either two CF_3-Tyr or CF_3-Phe residues provide insight into the structural perturbation elicited by the CF_3 groups and suggest that $^{TS}J_{FF}$ couplings are readily observable for ^{19}F–^{19}F distances that, in the time average, are greater than the sum of the van der Waals radii. The finding is underlined by the observation of ^{19}F–^{19}F TOCSY cross-peaks between two solvent-exposed TFA-Lys residues in the 27 kDa protein AncCDT-1[35] and the 25 kDa monomeric red fluorescent protein (mRFP).[36]

RESULTS

Site-specific incorporation of CF_3-Tyr and CF_3-Phe

The amino acids CF_3-Tyr and CF_3-Phe were incorporated into the protein PpiB by replacing Phe27 and Phe98 with the respective non-canonical amino acids, producing the samples PpiB CF_3-Tyr(27,98) and PpiB CF_3-Phe(27,98), respectively. As Phe4 is close to both residues in position 27 and 98,[37] and to provide more space for the CF_3 groups, two more samples were prepared that contained the mutation Phe4Ala in addition to the CF_3 residues in positions 27 and 98. In the following, we refer to these samples as PpiB F4A/CF_3-Tyr(27,98) and PpiB F4A/CF_3-Phe(27,98), respectively. Finally, a version of PpiB was produced with three CF_3-Tyr residues, yielding PpiB CF_3-Tyr(4,27,98). The mutant proteins were produced with the help of a previously published aminoacyl-tRNA synthetase, tfm-Phe2.[31] A recently published aminoacyl-tRNA synthetase selected for incorporation of SF_5-Phe, SF61,[30] proved to be poly-specific with regard to the incorporation of CF_3-Tyr and CF_3-Phe, but the incorporation efficiencies of this enzyme were not significantly enhanced over those achieved with tfm-Phe2 (Figure S1). Therefore, tfm-Phe2 was used for all subsequent work. The mutant samples of PpiB were obtained in good yield (typically 35 mg and 19 mg purified protein per liter of cell culture for the double and triple mutants, respectively) and purity as confirmed by mass spectrometry (Figures S2, S3 and Table S2). The 1D ^{19}F NMR spectra recorded on a 400 MHz NMR spectrometer (magnetic field strength 9.4 T) showed two well-separated lines with widths at half-height of 14 and 18 Hz for PpiB CF_3-Tyr(27,98) and PpiB CF_3-Phe(27,98), respectively (Figure 1). A corresponding spectrum recorded on a 700 MHz NMR spectrometer displayed a line width of 25 Hz for PpiB CF_3-Tyr(27,98), highlighting the impact of CSA relaxation on the ^{19}F NMR line width and sensitivity. Therefore, all subsequent spectra were recorded at 9.4 T. The spectra were recorded without ^{1}H decoupling as $^3J_{HF}$ and $^4J_{HF}$ couplings in phenyltrifluoromethane are only about 0.5 Hz[13] and CF_3-Tyr did not display resolved J_{HF} couplings either. Large changes in ^{19}F chemical shifts observed for the F4A mutants may be attributed to aromatic ring currents generated by Phe4 as well as the well-known sensitivity of ^{19}F nuclei with regard to changes in the chemical environment (Figure 1). In all spectra, each CF_3 group is characterized by a single chemical shift, indicating that rotation of the CF_3 groups is fast on the chemical shift time scale.

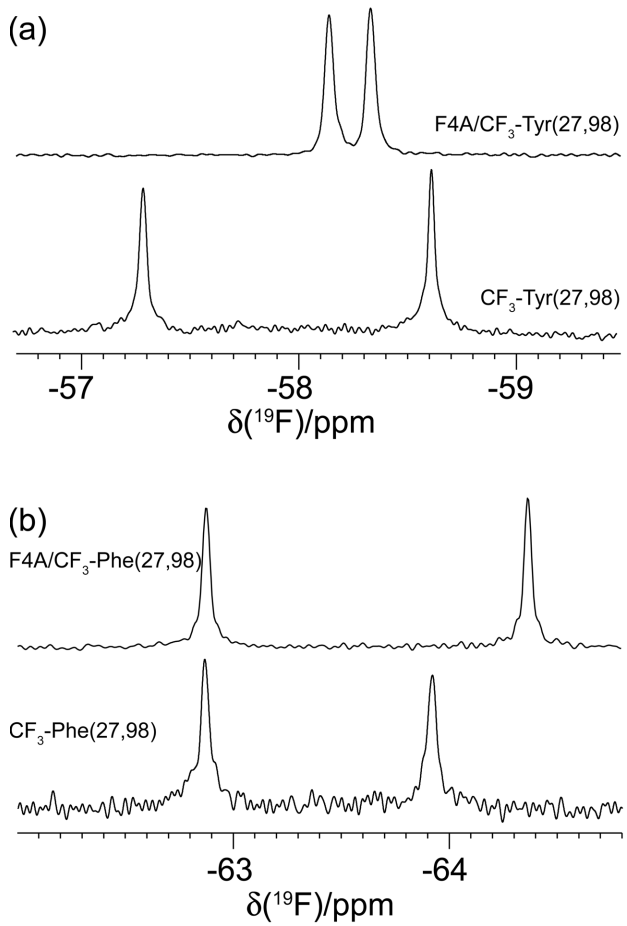

Figure 1. 1D ^{19}F NMR spectra of PpiB with two CF_3-Tyr or CF_3-Phe residues. Spectra recorded on a 400 MHz NMR spectrometer with an acquisition time of 67 ms. (a) Bottom panel: CF_3-Tyr in positions 27 and 98, PpiB CF_3-Tyr(27,98). Upper panel: sample with the additional mutation Phe4Ala. (b) Bottom panel: CF_3-Phe in positions 27 and 98, PpiB CF_3-Phe(27,98). Upper panel: sample with the additional mutation Phe4Ala.

Observation of through-space scalar J_{FF} couplings

To explore the presence of scalar through-space $^{TS}J_{FF}$ couplings, we recorded a ^{19}F–^{19}F DQF-COSY spectrum of PpiB CF_3-Tyr(27,98). Weak cross-peaks were observed but, despite the double-quantum filter, the diagonal peaks were considerably more intense and of mixed-phase appearance (Figure 2). Simulating the spectrum using the program Spinach[37] confirmed these observations and indicated that the mixed-phase appearance of the diagonal peaks can be attributed to cross-correlation between dipolar and CSA relaxation within each CF_3 group. While it proved difficult to produce a quantitative agreement between the experimental and simulated spectra without knowledge of the exact CSA tensors of the individual ^{19}F spins in a CF_3 group, the simulations confirmed that sizeable cross-peaks can be expected only if the two CF_3 groups are connected by scalar couplings.

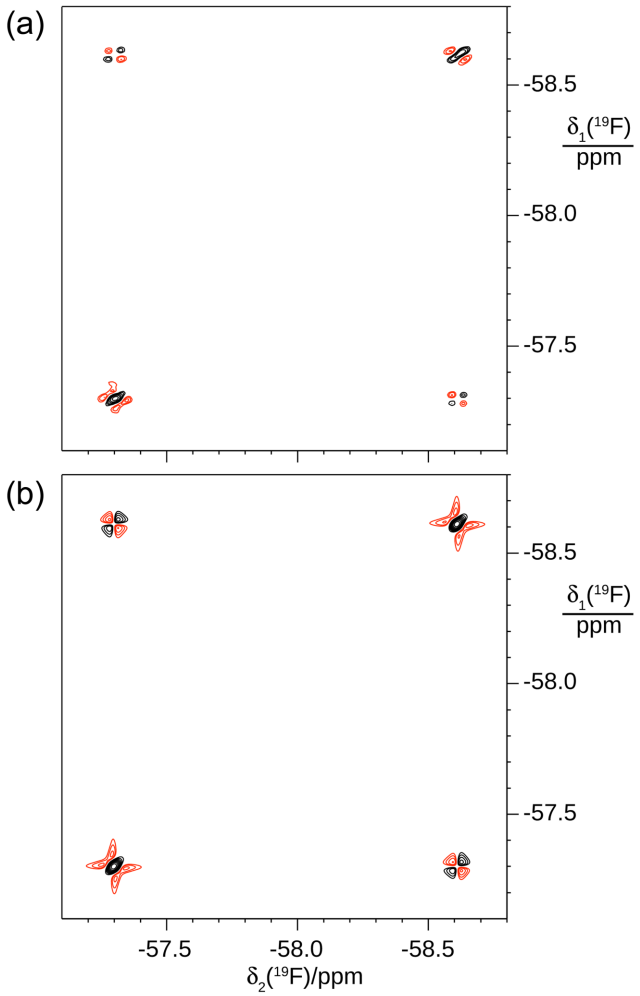

Figure 2. Experimental and simulated $^{19}F–^{19}F$ DQF-COSY spectra of 1 mM PpiB CF_3-Tyr(27,98). (a) Experimental spectrum recorded in 4 h, using t_{1max} = 64 ms and t_{2max} = 128 ms. (b) Spectrum simulated with a scalar coupling of J_{FF} = 2.5 Hz, using the program Spinach.[38] The simulation attributed an axially symmetric CSA tensor magnitude of Δ = 25 ppm to each CF_3 group along with dipolar $^{19}F–^{19}F$ relaxation.

$^{TS}J_{FF}$ couplings depend on orbital overlap and the ^{19}F spins involved may thus be expected to be sufficiently close to allow the observation of $^{19}F–^{19}F$ NOEs. A $^{19}F–^{19}F$ NOESY spectrum indeed revealed weak NOE cross-peaks (Figure 3a). As the chemical shifts of the ^{19}F NMR signals are well separated, the spectrum could be recorded with a very short acquisition time in the indirect dimension. This also made it much easier to probe for the existence of the $^{TS}J_{FF}$ couplings by $^{19}F–^{19}F$ TOCSY than COSY experiments, which need long evolution times to generate sizeable cross-peaks. $^{19}F–^{19}F$ TOCSY cross-peaks were readily observable within half an hour (Figure 3b). In a series of TOCSY spectra with different mixing times, the most intense cross-

peaks were observed for a mixing time of about 55 ms. The $^{19}F–^{19}F$ interactions were easier to observe in TOCSY than NOESY spectra, as illustrated by the cross-sections shown in Figure 3.

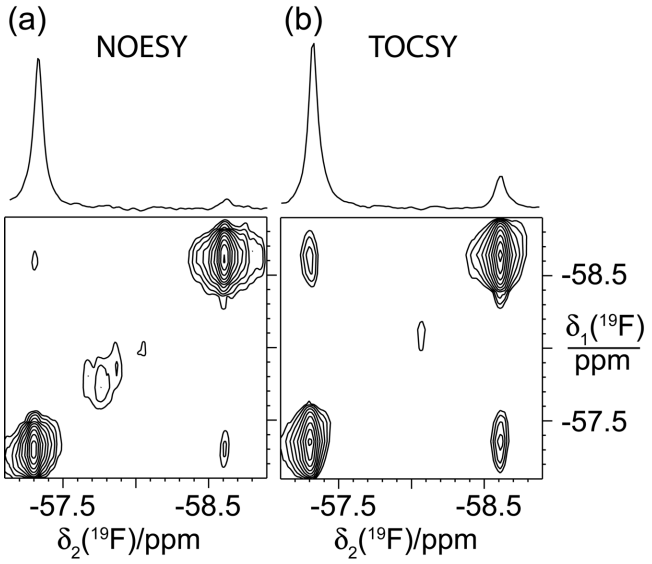

Figure 3. $^{19}F–^{19}F$ NOESY and $^{19}F–^{19}F$ TOCSY spectra of 0.6 mM PpiB CF_3-Tyr(27,98). Additional diagonal peaks of low intensity increased with time and indicate beginning sample degradation after two days at 25 °C. (a) NOESY spectrum recorded in 1 h with a mixing time of 200 ms and t_{1max} = 6 ms, t_{2max} = 80 ms. (b) TOCSY spectrum recorded in 0.5 h with 41 ms DIPSI-2 mixing, using t_{1max} = 5.1 ms, t_{2max} = 80 ms. 1D cross-sections through the low-field diagonal peaks in the respective spectra are shown in the panel above.

Structural relationship of CF_3 groups in PpiB CF_3-Tyr(27,98)

To probe the chemical environment of the CF_3 groups, we recorded a heteronuclear $^{1}H–^{19}F$ NOESY (HOESY) spectrum. In the crystal structure of the wild-type protein (PDB ID: 2NUL [37]), the side chain of Phe4 intercalates between the side chains of Phe27 and Phe98. If this structure is conserved in PpiB CF_3-Tyr(27,98), the CF_3 groups would be expected to be similarly separated by Phe4 and each CF_3 group would be near different methyl groups of leucine and isoleucine residues. The heteronuclear $^{1}H–^{19}F$ NOESY spectrum indeed displayed cross-peaks with protons in the aromatic and methyl regions (Figure 4). In the aromatic region, they appeared at the same ^{1}H chemical shift for both CF_3 groups, as expected for close proximity to Phe4, whereas they contacted different methyl groups, again as expected from the crystal

structure of the wild-type protein. $^{1}H–^{19}F$ NOESY experiments of proteins are low in sensitivity,[39] which is corroborated by the present results.

The $^{19}F–^{19}F$ NOESY spectrum showed a ratio of cross-peak to diagonal peak intensities of 1:24. For two ^{19}F spins rigidly positioned in a protein with 9 ns rotational correlation time, this suggests a $^{19}F–^{19}F$ distance of about 3.6 Å, which is greater than twice the van der Waals radius of fluorine (1.47 Å). As this estimate ignores the contributions from the additional ^{19}F spins in the CF_3 groups, which would be expected to enhance the NOE, it may be concluded that the actual minimal separation between the two different CF_3 groups is slightly longer. The cross-peak to diagonal peak ratio did not change, when the NOESY experiment was conducted with a ^{19}F inversion pulse in the middle of the mixing time to suppress spin-diffusion via nearby protons.[40]

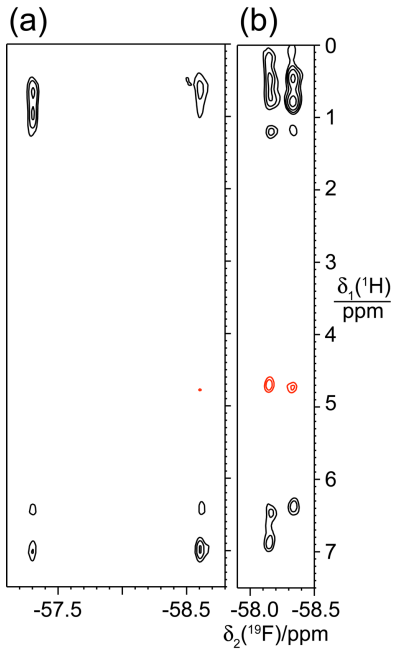

Figure 4. $^{1}H–^{19}F$ NOESY spectra of PpiB CF_3-Tyr(27,98) without and with the additional Phe4Ala mutation. The spectra were recorded with a mixing time of 200 ms, using t_{1max} = 7.3 ms and t_{2max} = 135 ms. Positive and negative peaks are plotted with black and red contour lines, respectively. Negative cross-peaks appear at the ^{1}H chemical shift of the water resonance. (a) 0.1 mM PpiB CF_3-Tyr(27,98). Spectrum recorded in 29 h. (b) 0.8 mM PpiB F4A/CF_3-Tyr(27,98). Spectrum recorded in 10 h.

To gain more insight into the distance between the CF_3 groups, we determined the crystal structure of PpiB CF_3-Tyr(27,98). Crystallographic data and refinement statistics are provided in Table S3. The crystal displayed two different conformations in the asymmetric unit. In the first structure (chain A in Figure 5), the

shortest inter-residual F–F distance is 5.6 Å. In the other structure (chain B in Figure 5), the side chain of Phe4 has moved aside and the CF_3 groups come as close as 2.8 Å of each other. This close contact explains the observation of scalar $^{TS}J_{FF}$ couplings, although the degree with which this conformation is populated in solution is unclear.

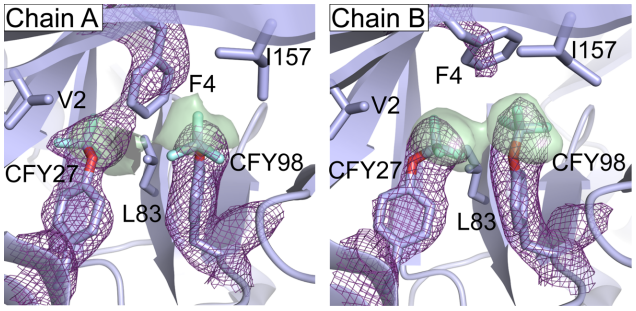

Figure 5. 2.0 Å crystal structure of PpiB CF_3-Tyr(27,98) (PDB ID: 7N3J), illustrating the positions of the CF_3 groups in the two conformations observed in the crystal (chains A and B). The CF_3O-groups are highlighted in cyan and red. The $2mF_O$-DF_C electron density map of the side chains of Phe4 and the CF_3-Tyr residues (labeled CFY27 and CFY98) is shown as a mesh contoured at 1.0 σ. The CF_3 groups were modeled into the mF_O-DF_C difference density map (generated through omitting the CF_3 groups from the model), indicated by the green surface contoured to 3 σ. Side chains of nearby hydrophobic residues are shown in a stick representation.

Phe4Ala mutant of PpiB CF_3-Tyr(27,98)

To eliminate interference from the side chain of Phe4, we produced the Phe4Ala mutant, PpiB F4A/CF_3-Tyr(27,98). Figure 6 compares ^{19}F–^{19}F NOESY and ^{19}F–^{19}F TOCSY spectra. Both the NOESY and TOCSY cross-peaks were more intense relative to the diagonal peaks than in PpiB CF_3-Tyr(27,98), indicating that the CF_3 groups had moved closer to each other and the scalar $^{TS}J_{FF}$ coupling between them had increased, suggesting a greater population of the conformation B. The ^{1}H–^{19}F NOESY spectrum showed no evidence of cross-peaks to ^{1}H spins common to both CF_3 groups.

Interestingly, negative cross-peaks appeared at the ^{1}H NMR frequency of the water resonance, indicating the presence of rapidly moving water molecules in the cavity created by the Phe4Ala mutation (Figure 4B). Negative NOESY cross-peaks are expected when the correlation time τ_c of the dipolar interaction between the ^{19}F spins of the CF_3 group and the ^{1}H spins of the water molecules is short. For a 2-spin system comprising one ^{1}H and one ^{19}F spin, the common equation for the NOE[41] predicts the sign change for τ_c = 0.46 ns on a 400 MHz NMR spectrometer.

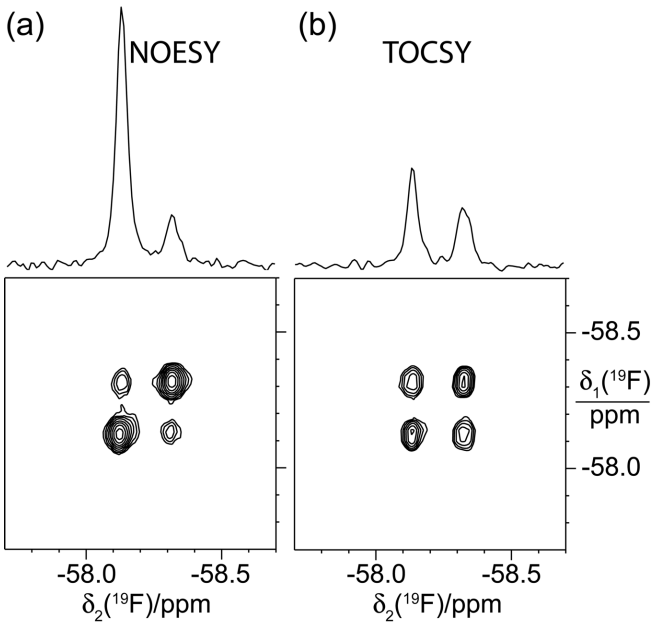

Figure 6. $^{19}F–^{19}F$ NOESY and $^{19}F–^{19}F$ TOCSY spectra of 0.8 mM PpiB F4A/CF_3-Tyr(27,98). The spectra were recorded using t_{1max} = 4 ms and t_{2max} = 135 ms. (a) NOESY spectrum recorded in 1 h with a mixing time of 200 ms. (b) TOCSY spectrum recorded in 0.5 h with 41 ms DIPSI-2 mixing, using half the number of scans per FID compared to the NOESY spectrum. 1D cross-sections through the low-field diagonal peaks are shown in the panel above.

Quantitative measurements of $^{TS}J_{FF}$ couplings

To measure the magnitude of the $^{TS}J_{FF}$ couplings more quantitatively, we designed a quantitative *J* evolution experiment that determines the amount of in-phase magnetization remaining after a spin-echo delay implemented either with or without inversion of the coupling partner in the middle of the delay (Figure 7). The pulse sequence is based on an experiment reported previously for measuring small $^1H–^{113}Cd$ coupling constants[24] and was adapted for the homonuclear case of two resonances by using jump-return elements[42] to achieve selective excitation and inversion in the shortest time possible. As the CF_3 groups rotate, the $^{TS}J_{FF}$ couplings are averaged to a single value between all inter-CF_3 couplings, leading to an attenuation of the in-phase product operator term by $\cos^3\pi J\tau$, where *J* is the coupling constant and τ the total spin-echo delay. The experiment included a 90° pulse at the end of the spin-echo delay to select only in-phase magnetization for detection. Fitting the data indicated $^{TS}J_{FF}$ couplings of 2.5 Hz for PpiB CF_3-Tyr(27,98) and a larger coupling (about 5.8 Hz) for PpiB F4A/CF_3-Tyr(27,98), in agreement with the larger $^{19}F–^{19}F$ TOCSY cross-peaks observed in the latter sample (Figure 8). A control experiment using the immunoglobulin-binding domain of protein G, GB1, with Trp43 replaced by 7-fluorotryptophan confirmed that the experiment reliably indicates the absence of J_{FF} couplings in a protein containing only a single ^{19}F spin (Figure S5).

We hypothesized that the larger $^{TS}J_{FF}$ coupling might be observable in the multiplet fine structures of the 1D ^{19}F NMR spectrum of PpiB F4A/CF_3-Tyr(27,98), if the FID is processed with strong resolution enhancement. Although we could not fully resolve the multiplet fine structures, the spectrum indicates the 1:3:3:1 quartet fine structure for each of the CF_3 groups, including the expected roof effect and a splitting in agreement with the coupling constant measured by the quantitative *J* evolution experiment (Figure 9).

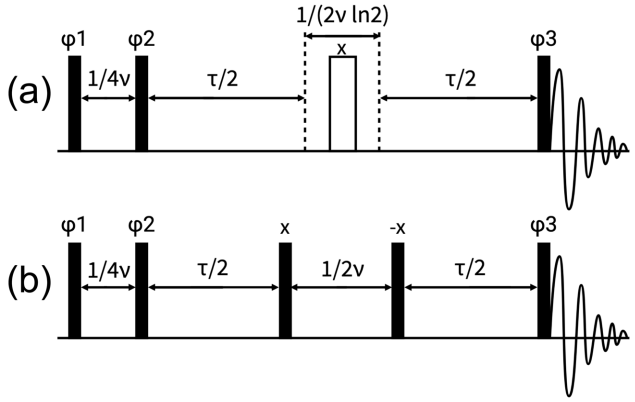

Figure 7. Pulse sequence for the quantitative *J* evolution experiment for measuring homonuclear coupling constants in an NMR spectrum with two resonances. Solid and open bars represent 90° and 180° pulses, respectively. ν denotes the frequency difference in Hz between the resonances of the coupling partners. The carrier frequency is set to the frequency of one of the resonances. The pulses with phases ϕ_1 and ϕ_2 form a jump-return element[42] with 90° excitation of the coupling partner and the pulses separated by 1/2ν form a 180° jump-return element. τ is the spin-echo delay for coupling evolution. Phase cycle: ϕ_1={x,-x}, ϕ_2={-x,x}, ϕ_3={x,x,-x,-x}, and ϕ_{rec}={x,-x,x,-x}. (a) Experiment with active coupling evolution during the spin-echo delay. The delay 1/[2ν ln(2)] aims to account for the additional relaxation occurring in the reference experiment during the 180° jump-return element. (b) Reference experiment with refocusing of the $^{TS}J_{FF}$ couplings. Pulsed field gradients (not shown) were applied at the start and end of the τ delay in both experiments with a pulse length of 1 ms and field strength of 15 G cm^{-1}.

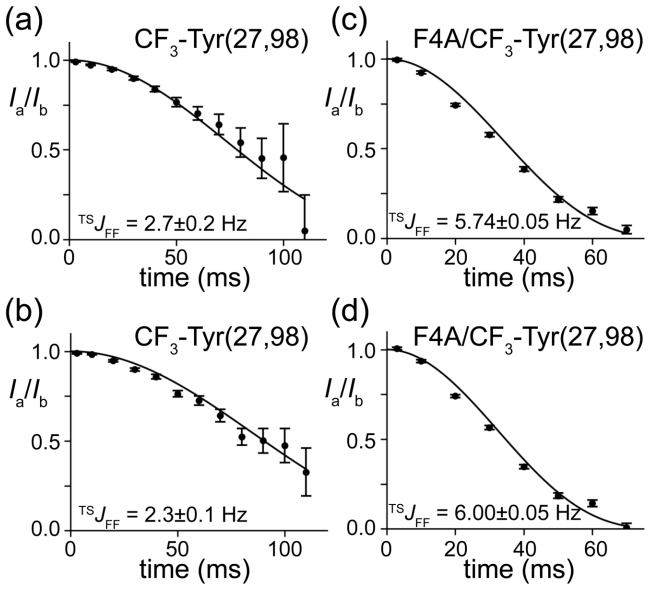

Figure 8. Relative peak intensities observed in a quantitative *J* evolution experiment conducted with and without refocusing of *J* couplings during the spin-echo delay. The relative peak intensities observed in experiments a and b of Figure 7 are plotted versus the spin echo delay τ. The $^{TS}J_{FF}$ coupling was determined by fitting the function $\cos^3 \pi J t$. The uncertainty bars were derived from the level of white noise. Uncertainties in fitted coupling constants were estimated by Monte Carlo simulation with 1000 iterations. For each iteration, noise sampled from the spectrum was added to the peak intensities and a coupling constant was fitted. The final uncertainty was captured by the standard deviation in fitted coupling constants over all iterations. The best fitting $^{TS}J_{FF}$ values are given in each plot. (a) Signal at -57.5 ppm in PpiB CF_3-Tyr(27,98). (b) Same as (a), but for the peak at -58.8 ppm. (c) Signal at -58.15 ppm in PpiB F4A/CF_3-Tyr(27,98). The small difference in chemical shifts between the ^{19}F NMR signals in this mutant resulted in an intensity ratio of 0.93 instead of 1.0 for the first point. Consequently, all points were divided by 0.93 prior to fitting. (d) Same as (c), but for the peak at -58.35 ppm.

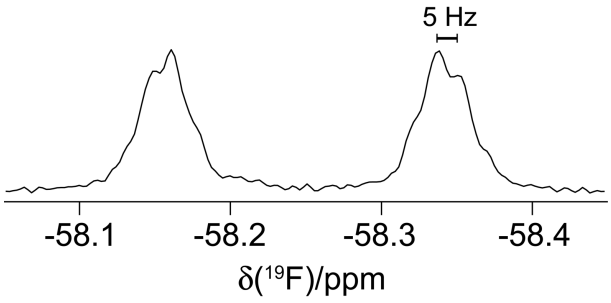

Figure 9. 1D ^{19}F NMR spectrum of PpiB F4A/CF_3-Tyr(27,98) processed with Lorentz-Gauss multiplication for resolution enhancement to visualize the 1:3:3:1 multiplet fine structure. The quantitative *J* evolution experiment measured the $^{TS}J_{FF}$ coupling as close to 6 Hz. The scale bar corresponds to 5 Hz.

Larger $^{TS}J_{FF}$ couplings are expected to facilitate the observation of ^{19}F–^{19}F DQF-COSY cross-peaks. Indeed, the spectrum recorded of PpiB F4A/CF_3-Tyr(27,98) showed a simpler appearance than that of PpiB CF_3-Tyr(27,98), with less dominant effects from CSA–DD cross-correlated relaxation (Figure 10).

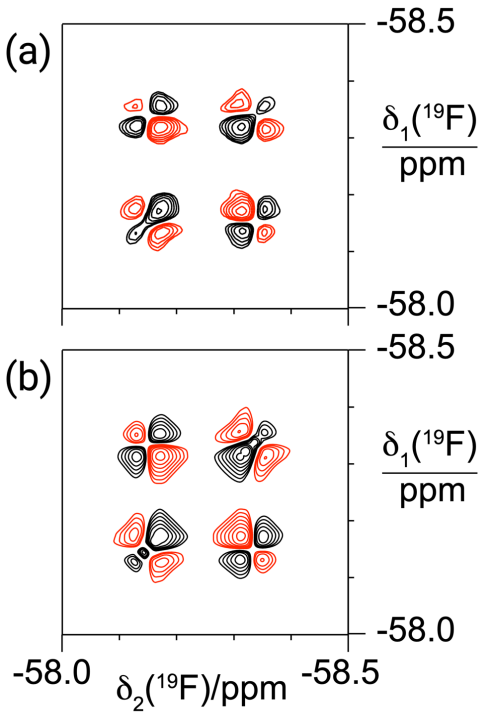

Figure 10. Experimental and simulated $^{19}F–^{19}F$ DQF-COSY spectra of 0.8 mM PpiB F4A/CF_3-Tyr(27,98). (a) Experimental spectrum recorded in 2.7 h, using t_{1max} = 68 ms and t_{2max} = 135 ms. (b) Spectrum simulated with a scalar coupling of J_{FF} = 6.0 Hz, using the program Spinach[38] and the same CSA parameters as in Figure 2.

CF_3-Phe mutants

Instead of providing more space for the OCF_3 groups by mutating Phe4 to alanine, we also explored reducing the space demands of the non-canonical amino acids by making CF_3-Phe mutants. Figure 10 shows the $^{19}F–^{19}F$ NOESY and TOCSY spectra of PpiB CF_3-Phe(27,98). The spectra confirmed the close proximity of the CF_3 groups and presence of $^{TS}J_{FF}$ couplings (Figure 11), but the sample was prone to precipitation. Attributing the lesser stability of the protein to the less flexible positioning of CF_3 compared with OCF_3 groups, we also produced the mutant PpiB F4A/CF_3-Phe(27,98). This mutant proved to be much more stable against precipitation. The appearance of the $^{19}F–^{19}F$ NOESY and TOCSY spectra was similar to those obtained for PpiB CF_3-Phe(27,98) (Figure S4). The quantitative *J* evolution experiment indicated a $^{TS}J_{FF}$ coupling of about 3.5 Hz for both samples (Figure S6).

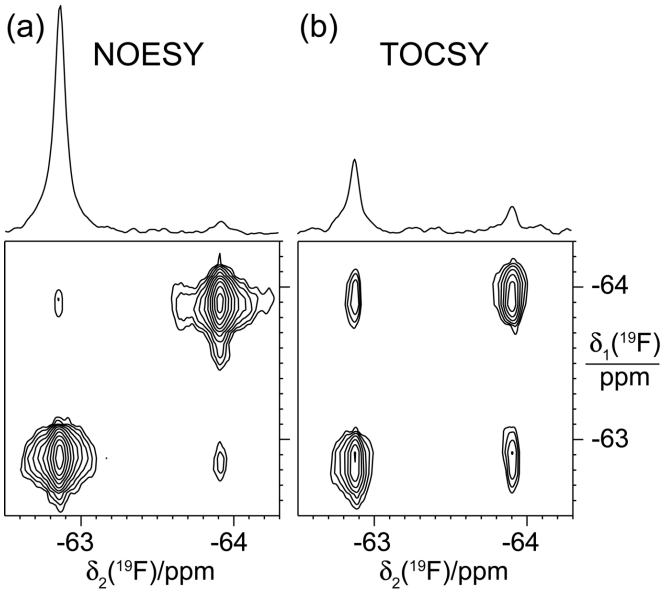

Figure 11. ^{19}F–^{19}F NOESY and TOCSY spectra of 0.5 mM PpiB CF_3-Phe(27,98). The spectra were recorded using t_{1max} = 6 ms and t_{2max} = 135 ms. (a) NOESY spectrum recorded in 7 h with a mixing time of 200 ms. (b) TOCSY spectrum recorded in 50 minutes with 41 ms DIPSI-2 mixing, using 8 times fewer scans per FID than in the NOESY spectrum. The top panel shows cross-sections through the low-field diagonal peak.

Assuming that the PpiB F4A/CF_3-Phe(27,98) construct would populate a single conformation, we determined its single-crystal X-ray structure. Crystallographic data and refinement statistics are provided in Table S3. The asymmetric unit indeed revealed a single conformation (Figure 12), with the shortest F–F distance between the CF_3 groups being 3.4 Å. This is more than twice the van der Waals radius of fluorine (1.47 Å), suggesting that observable $^{TS}J_{FF}$ couplings either require little inter-residual overlap of fluorine orbitals or transient F–F contacts occur due to rotation of the CF_3 groups or conformational variability of the protein in solution.

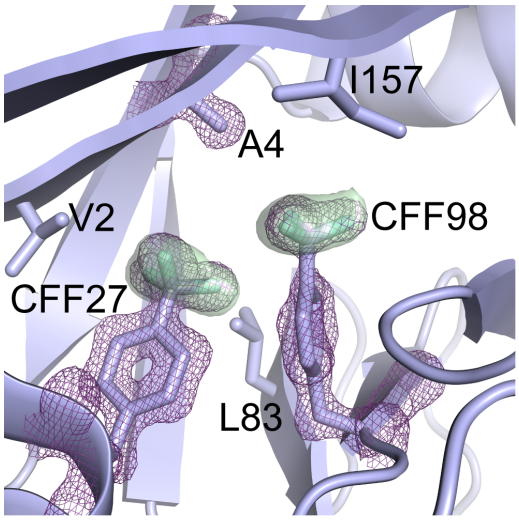

Figure 12. 1.35 Å crystal structure of PpiB F4A/CF_3-Phe(27,98) (PDB ID: 7RFD), illustrating the positions of the CF_3 groups, which are highlighted in cyan. The 2mF_O-DF_C electron density map of the side chains of Ala4 and the CF_3-Phe residues (labeled CFF27 and CFF98) is shown as a mesh contoured at 1.2 σ. The CF_3 groups were modeled into the mF_O-DF_C density (generated by omitting the CF_3 groups from the model), indicated by the green surface contoured to 3.5 σ. Side chains of nearby hydrophobic residues are shown in a stick representation.

A ^{1}H–^{19}F NOESY spectrum of PpiB F4A/CF_3-Phe(27,98) showed negative NOE cross-peaks at the ^{1}H chemical shift of the water resonance, confirming the occurrence of rapidly moving water molecules in the cavity produced by the Phe4Ala mutation (Figure S8). The crystal structure showed no distinct electron density for these water molecules, as expected for highly disordered water.

PpiB CF_3-Tyr(4,27,98)

The crystal structure of wild-type PpiB contains voids in the hydrophobic core of the protein large enough to house water molecules although they appear empty in the crystal structure. This explains why the additional OCF_3 groups can be accommodated in PpiB CF_3-Tyr(27,98) without major structural adjustments of the polypeptide backbone. To explore the limits of this structural accommodation, we also substituted Phe4 by CF_3-Tyr in the triple-mutant PpiB CF_3-Tyr(4,27,98). The 1D ^{19}F NMR spectrum of this mutant showed three main peaks as expected. Two of them were at very similar chemical shifts as in the double-mutant PpiB CF_3-Tyr(27,98), assigning the new resonance at -55.9 ppm to CF_3-Tyr4. A ^{19}F–^{19}F TOCSY spectrum recorded with 41 ms mixing time showed cross-peaks between all three ^{19}F NMR signals, indicating $^{TS}J_{FF}$ couplings between all three spins and providing independent evidence that the CF_3 groups of residues 27 and 98 are in similar proximity of residue 4 (Figure 13). As the cross-peaks with CF_3-Tyr4 are three to

four times smaller than the cross-peaks between residues 27 and 98, the $^{TS}J_{FF}$ coupling constants of CF_3-Tyr4 are probably less than 2 Hz.

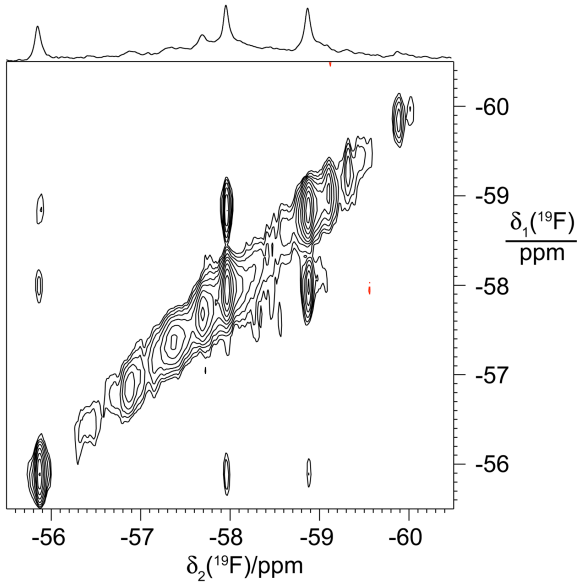

Figure 13. $^{19}F–^{19}F$ TOCSY spectrum of a 0.5 mM solution of PpiB CF_3-Tyr(4,27,98). Mixing time 41 ms, t_{1max} = 5.1 ms, t_{2max} = 82 ms, total recording time 24 h. The 1D ^{19}F NMR spectrum displayed at the top was recorded of the fresh sample prior to the 2D experiment.

Additional signals in the 1D ^{19}F NMR spectrum of PpiB CF_3-Tyr(4,27,98) (Figure 13) indicated the presence of alternative conformations, which increased in intensity during the NMR measurements. Furthermore, the protein precipitated heavily within a day. A second freshly prepared sample reproduced the heterogeneous 1D ^{19}F NMR spectrum. A $^{19}F–^{19}F$ NOESY spectrum (mixing time 200 ms) recorded with this sample yielded barely detectable NOE cross-peaks between residues 27 and 98, and the sensitivity was insufficient to observe cross-peaks with residue 4. This result shows that $^{TS}J_{FF}$ interactions between CF_3 groups can be easier to detect than $^{19}F–^{19}F$ NOEs.

Optimal $^{19}F–^{19}F$ TOCSY mixing times

The small magnitude of the $^{TS}J_{FF}$ couplings explains why peak splittings are generally not resolved in the 1D ^{19}F NMR spectra of proteins containing CF_3 groups. As $^{19}F–^{19}F$ TOCSY proved the most sensitive experiment for confirming the interaction between the CF_3 groups, we sought to identify the mixing time for most sensitive detection of the TOCSY cross-peaks. Simulations of DIPSI-2 mixing with the program

Spinach[36] suggested an optimal mixing time of 45 ms for PpiB CF_3-Tyr(27,98). Experimentally, a series of TOCSY spectra recorded with different mixing times showed a good match with the prediction (Figure S9). Simulations performed without relaxation indicate that achieving complete transfer of magnetization between CF_3 groups by TOCSY mixing is not as easy as between single ^{19}F spins (Figure S10).

^{19}F–^{19}F TOCSY cross-peaks between trifluoroacetyl-L-lysine residues

To demonstrate the potential of ^{19}F–^{19}F TOCSY for the detection of contacts between CF_3 groups installed in aliphatic amino acids, we incorporated two N^6-(trifluoroacetyl)-L-lysine (TFA-Lys) residues (Figure 14a) into the protein AncCDT-1. The structure of this protein comprises two domains with a binding site for a basic amino acid (arginine or lysine) between them. The crystal structure in the bound state (PDB ID 5T0W)[35] positions the C^α atoms of residues 68 and 162 within 12.5 Å (Figure 14b). Both residues are solvent exposed and we hypothesized that the CF_3-groups of TFA-Lys residues incorporated in positions 68 and 162 can form contacts across the domain interface, which may be manifested by cross-peaks in ^{19}F–^{19}F NOESY or TOCSY spectra.

An aminoacyl-tRNA synthetase for genetic encoding of TFA-Lys has been published previously,[43] but the protein expression yields reported were low. We therefore selected a new synthetase for TFA-Lys from our library based on the pyrrolysyl-tRNA synthetase derived from the methanogenic archaeon ISO4-G1,[44] using a recently established fluorescence activated cell sorting (FACS) protocol (see SI Section 3 for details).[45] The enzyme identified from this selection incorporates TFA-Lys with high specificity in response to amber stop codons, provided that the amino acid is present in high concentration. The ^{19}F NMR spectrum of AncCDT-1 with Pro68 and Asn162 mutated to TFA-Lys, in the following referred to as AncCDT-1 TFA-Lys(68,162), displayed two signals separated by 0.17 ppm. Indeed, a ^{19}F–^{19}F TOCSY spectrum displayed clear cross-peaks, whereas we were unsuccessful to detect the interaction in a ^{19}F–^{19}F NOESY experiment (Figure 14, c and d). AncCDT-1 has a greater molecular weight than PpiB (27 versus 19 kDa), yet the line widths of the ^{19}F NMR signals of AncCDT-1 TFA-Lys(68,162) are much narrower than those of the CF_3 groups in the PpiB mutants of Figure 1 (about 7 Hz versus 14–17 Hz), indicating significantly increased mobility of CF_3 groups at the end of a solvent-exposed aliphatic amino acid versus the CF_3 groups of buried aromatic residues. Remarkably, the mobility of the TFA-Lys side chains did not prevent the buildup of readily observable inter-residual TOCSY cross-peaks via $^{TS}J_{FF}$ interactions.

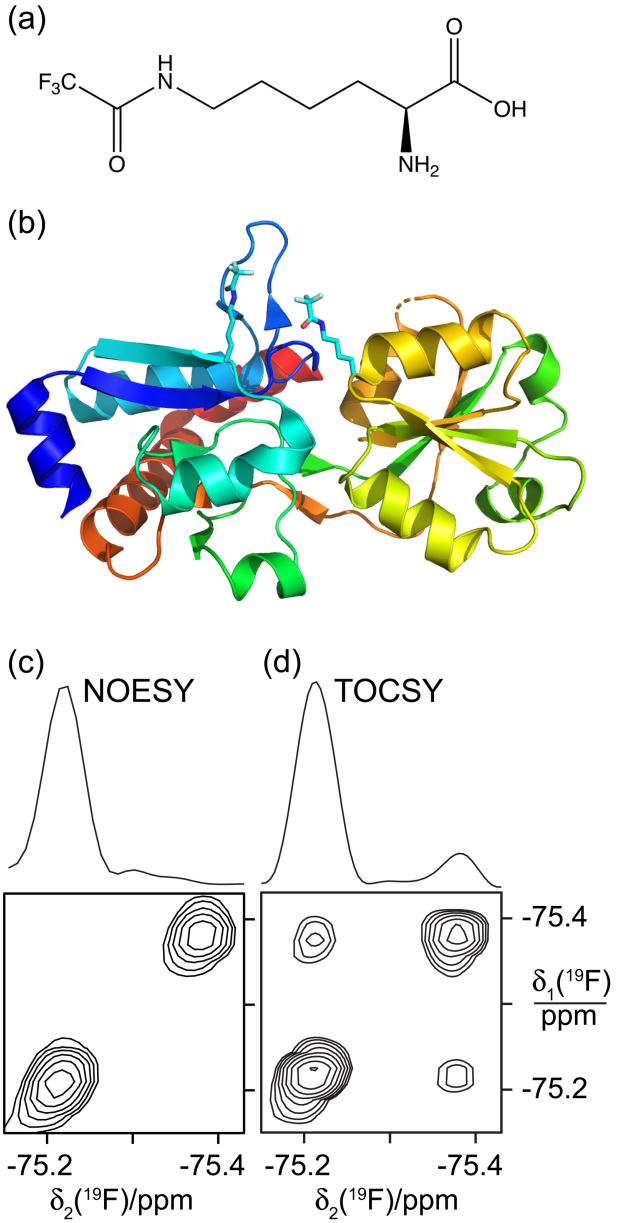

Figure 14. Fluorine–fluorine contacts between solvent-exposed CF_3 groups detected by ^{19}F–^{19}F TOCSY. (a) Chemical structure of N^6-(trifluoroacetyl)-L-lysine (TFA-Lys). In a fully extended conformation, the distance between the C^α atom and the CF_3 group is about 9.2 Å. (b) Cartoon representation of AncCDT-1 (crystal structure 5T0W[35]) with TFA-Lys residues modeled at sites 68 and 162 (numbering as in the crystal structure) shown in a stick representation. (c) ^{19}F–^{19}F NOESY spectrum recorded of 150 µM AncCDT-1 TFA-Lys(68,162) in 24 h with a mixing time of 200 ms, using t_{1max} = 40 ms and t_{2max} = 68 ms. The cross-section through the low-field diagonal peak is shown above the 2D spectrum. (d) ^{19}F–^{19}F TOCSY spectrum recorded in 9.5 h with 100 ms DIPSI-2 mixing. All other parameters as in (c).

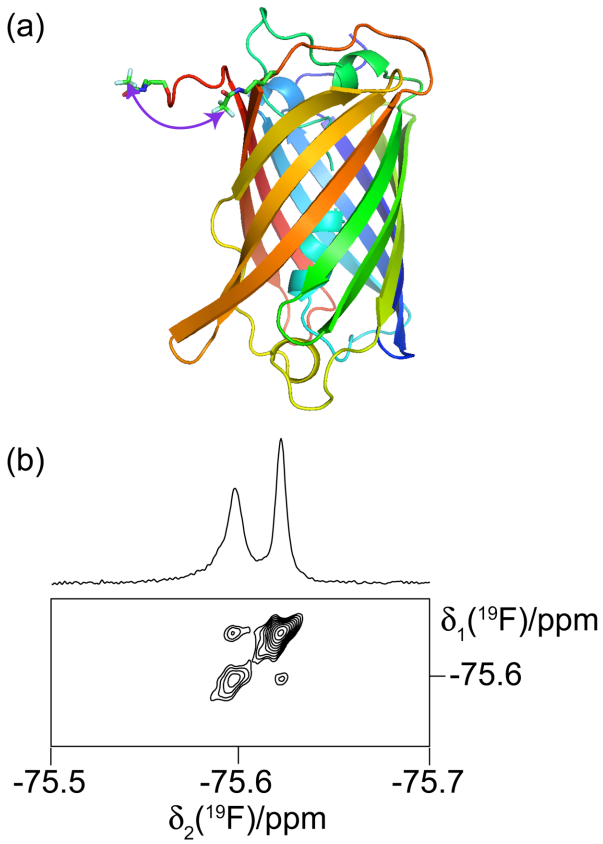

Figure 15. Transient fluorine–fluorine contacts between CF_3 groups detected by ^{19}F–^{19}F TOCSY. (a) Crystal structure 2QLG[46] of mRFP (cartoon drawing) with TFA-Lys residues (stick representation) modeled in position 192 and at the C-terminus (position 225). The double-headed arrow points at the CF_3 groups. (b) ^{19}F–^{19}F TOCSY spectrum recorded of a 1.6 mM solution of the protein, using a mixing time of 400 ms and a total recording time of 25 minutes. The 1D ^{19}F NMR spectrum is plotted above.

To further test the scope of detecting small $^{TS}J_{FF}$ couplings between flexible TFA-Lys residues, we installed a TFA-Lys residue in mRFP at position 192 and a second TFA-Lys residue at the C-terminus at position 225 (Figure 15a). Despite the high mobility of both TFA-Lys residues (there is no electron density for the C-terminal amino acid, Ala225, in the crystal structure), their respective ^{19}F NMR resonances could be resolved in the 1D ^{19}F NMR spectrum with a chemical shift difference of 0.025 ppm and a linewidth at half height of about 3.5 and 2 Hz. The slow relaxation rates allowed recording a ^{19}F–^{19}F TOCSY spectrum with a very long mixing time, which revealed a cross-peak (Figure 15b). In this way the experiment confirmed that the C-terminus of mRFP is in general proximity of the position 192.

DISCUSSION

The wide-spread adoption of high-field NMR spectrometers in the past decades has made the detection of through-space scalar $^{19}F–^{19}F$ couplings more challenging as the line widths in the ^{19}F NMR spectrum increase with increasing magnetic field strength due to CSA relaxation. On our 9.7 T NMR instrument, the full line widths at half height (14 Hz and 17 Hz for the PpiB CF_3-Tyr and PpiB CF_3-Phe samples, respectively) at 25 °C compared favorably with the line width observed for the SF_5 groups in PpiB samples prepared with SF_5-Phe in positions 27 and 98 (about 50 Hz),[30] but were still three to five-fold larger than the $^{TS}J_{FF}$ couplings observed between the CF_3 groups. Nonetheless, detection of these couplings proved straightforward by $^{19}F–^{19}F$ TOCSY. Detection was aided by the absence of ^{19}F NMR signals from the canonical amino acids in the protein sample, the capacity to produce protein samples with non-canonical fluorinated amino acids with high yield and specificity, and the small size of J_{HF} couplings in CF_3-Tyr and CF_3-Phe (about 0.5 Hz),[13] allowing us to record all data without ^1H decoupling. In addition, the rotation of the CF_3 groups is fast on the chemical shift time scale in all our samples, leading to degenerate ^{19}F chemical shifts of each CF_3 group. With each diagonal peak representing three equivalent ^{19}F nuclei, a cross-peak between two CF_3 groups can be expected to be up to nine times more sensitive than a corresponding cross-peak between single ^{19}F spins, but the peak heights are attenuated by any multiplet splittings arising from $^{TS}J_{FF}$ couplings. When linked by a $^{TS}J_{FF}$ coupling, the CF_3 groups are expected to display quartets, which we were able to trace in the sample with the largest $^{TS}J_{FF}$ coupling, PpiB F4A/CF_3-Tyr(27,98) (Figure 9).

Compared to a two-spin system, simulations indicated that the complexity of the multiplet fine structures hinders complete magnetization transfer in TOCSY spectra and shifts the optimal magnetization transfer to shorter mixing times. Nonetheless and unexpectedly, detecting the fluorine–fluorine interactions by $^{19}F–^{19}F$ TOCSY compared favorably with their detection by $^{19}F–^{19}F$ NOESY spectra.

The high-resolution crystal structure of PpiB F4A/CF_3-Phe(27,98) (Figure 12) suggests that $^{TS}J_{FF}$ couplings can be measured also for F–F distances greater than the sum of the fluorine van der Waals radii. Previous studies of small synthetic molecules reported the detection of $^{TS}J_{FF}$ couplings for F–F distances up to almost 3.2 Å[47] combined with a steep exponential distance dependence, where the coupling constant doubles for every 0.14–0.21 Å decrease in F–F distance.[47,48] $^{TS}J_{FF}$ couplings are also sensitive to the relative spatial orientation of the C-F bonds, as $^{TS}J_{FF}$ couplings of 7.9 Hz and 1.9 Hz have been reported for F–F distances of 3.18 Å and 3.02 Å, respectively.[47] In PpiB F4A/CF_3-Phe(27,98), $^{TS}J_{FF}$ = 3.5 Hz observed between the CF_3 groups appears large, considering that the value presents an average of all F–F interactions and only

the shortest corresponds to a F–F distance of 3.4 Å. This suggests that the $^{TS}J_{FF}$ coupling is greatly amplified by the dynamics of the protein, with transient short F–F distances between the CF_3 groups contributing disproportionately. Combined with geometric dependences on the relative orientation of molecular orbitals,[17,48] it is difficult to convert $^{TS}J_{FF}$ couplings into quantitative distance restraints beyond an upper limit near the van der Waals distance.

In the absence of sizeable J_{HF} couplings, correlating the ^{19}F and ^{1}H NMR spectra can be achieved by heteronuclear ^{1}H–^{19}F NOESY spectra, but these are fairly insensitive. The far lesser sensitivity of the heteronuclear experiment compared with the ^{19}F–^{19}F NOESY spectrum is expected, as the heteronuclear ^{1}H–^{19}F NOE in slowly tumbling macromolecules depends mostly on the spectral density at the difference in Larmor frequencies of ^{1}H and ^{19}F spins rather than the much greater spectral density at zero frequency, which dominates the homonuclear ^{19}F–^{19}F NOE.[4] In principle, ^{1}H–^{19}F NOEs provide a way of assigning the ^{19}F NMR resonances if the ^{1}H NMR spectrum has been assigned. As the introduction of non-canonical amino acids changes the ^{1}H chemical shifts of the protein, however, it is advantageous to circumvent the assignment problem by installing a pair of non-canonical amino acids at specific sites, which can be selected by genetic encoding. In the case of the triple mutant PpiB CF_3-Tyr(4,27,98), the ^{19}F resonance of residue 4 was readily assigned by comparison with the spectrum of PpiB CF_3-Tyr(27,98) (Figures 3b and 13).

The crystal structure of PpiB CF_3-Tyr(27,98) suggests that the side chain of Phe4 can be positioned between the side chains of residues 27 and 98 like in the wild-type protein,[37] preventing direct contacts between the CF_3 groups. It is intriguing to speculate that the situation could be equivalent to that reported previously for 1,5,8-trifluoro-9,10-diphenylanthracene, where an intervening phenyl ring was found to greatly increase the $^{TS}J_{FF}$ coupling between the fluorines in positions 1 and 8.[49] As the alternative conformation B in the crystal structure showed a direct F–F contact, however, a more conservative interpretation is to assume that the experimentally observed $^{TS}J_{FF}$ coupling is entirely due to direct contacts in conformation B rather than being mediated by the aromatic ring of Phe4. Also the high-resolution crystal structure of PpiB F4A/CF_3-Phe(27,98) suggests that the observation of $^{TS}J_{FF}$ couplings does not depend on stable direct fluorine–fluorine van der Waals contacts.

The increased space demand of the CF_3 variants of tyrosine and phenylalanine explains why samples of PpiB CF_3-Tyr(27,98), PpiB CF_3-Phe(27,98) and, most of all, PpiB CF_3-Tyr(4,27,98) precipitated during NMR measurements at 25 °C. We have shown previously that the β-barrel fold of PpiB can accommodate

SF_5 groups in the hydrophobic core of PpiB SF_5-Phe(27,98),[30] but also those samples were prone to precipitation. Nonetheless, the backbone structure of PpiB remained practically unchanged upon the introduction of additional OCF_3 groups. As the side chains of phenylalanine and tyrosine residues have a high propensity of co-localization in protein structures, $^{TS}J_{FF}$ couplings between site-specifically installed CF_3-Phe or CF_3-Tyr residues open a practical route to experimental verification of short-range contacts predicted by protein structures modeled by programs such as AlphaFold[49] and RoseTTAFold.[50]

In the present work, precipitation was successfully suppressed by introducing the additional mutation Phe4Ala, which makes room for the CF_3 groups in the hydrophobic core. Interestingly, these mutants displayed intense intermolecular ^1H–^{19}F NOEs between CF_3 groups and water molecules (Figures 4b and S8). The negative sign of these NOE cross-peaks indicates rapid reorientation of the vector between the ^{19}F spins of the CF_3 groups and the ^1H spins of water molecules on a time scale shorter than about 0.4 ns, indicating that the hydration of the CF_3 groups is kinetically labile. Rapid rotation of the CF_3 groups alone is not sufficient to change the sign of ^1H–^{19}F NOEs, as demonstrated by the spectra of Figures 4 and S8. To explain the negative sign of the NOE cross-peaks with water, the hydration water molecules must thus be subject to fast translational diffusion like in the common situation of negative ^1H–^1H NOE cross-peaks between surface hydration water and protein protons.[51] We believe this is the first time that hydration water molecules with such short residence times have been detected in a hydrophobic protein cavity.

The facile observation of $^{TS}J_{FF}$ couplings between two TFA-Lys residues in the protein AncCDT-1 is remarkable in view of the molecular weight of the protein (27 kDa) and the high flexibility of a solvent-exposed TFA-Lys side chain, which is manifested in relatively narrow ^{19}F NMR line widths. The narrow line widths suggest that ^{19}F–^{19}F contacts between two TFA-Lys residues are only transient, reducing the magnitude of the effective $^{TS}J_{FF}$ couplings and disfavoring magnetization transfer by ^{19}F–^{19}F NOEs. On the other hand, the relatively long transverse relaxation times enable the use of longer TOCSY mixing times and the detection of small $^{TS}J_{FF}$ couplings.

In previous work, we installed two N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine residues at the same sites in AncCDT-1 to observe ^1H–^1H NOEs. In this case, the ^1H NMR signals of the trimethyl silyl groups appeared at the same chemical shift and we had to incorporate two different non-canonical amino acids to observe an inter-residual NOE between them.[52] The pronounced sensitivity of ^{19}F chemical shifts to the local

chemical environment is illustrated by the chemical shift difference of almost 0.2 ppm observed in AncCDT-1 TFA-Lys(68,162).

In our mRFP mutant containing two TFA-Lys residues, only one was located in the folded part of the protein while the other was appended at the C-terminus. Despite the large conformational space available for these residues, an interaction between the CF_3 groups could readily be documented by a $^{19}F–^{19}F$ TOCSY cross-peak, demonstrating the potential of $^{TS}J_{FF}$ couplings for the detection of general proximity of solvent exposed sites. The interaction may be promoted by the tendency of perfluorocarbons to segregate from water or hydrocarbons.[53]

The new aminoacyl-tRNA synthetase for genetic encoding of TFA-Lys was derived from a pyrrolysyl-tRNA synthetase library based on the methanogenic archaeon ISO4-G1, that is more readily expressed in active form than the previously published synthetase derived from *Methanosarcina barkeri*.[43] Even though the new synthetase requires high concentrations of TFA-Lys to suppress the misincorporation of glutamine, the system is affordable due to the low cost of the amino acid. Our system thus presents an attractive alternative to chemical CF_3 tags, which have become increasingly popular for monitoring the response of proteins to ligand binding, such as drugs binding to G-protein coupled receptors, but mostly depend on single solvent-exposed cysteine residues to achieve site-selectivity.[54,55]

CONCLUSIONS

To the best of our knowledge, through-space $^{19}F–^{19}F$ couplings have not been reported for proteins since their first observation in 1978.[12] In our protein samples, the scalar couplings between CF_3 groups proved to be observable in $^{19}F–^{19}F$ TOCSY spectra with remarkable sensitivity on a two-channel 400 MHz NMR spectrometer, despite their magnitude being much smaller than the ^{19}F NMR line widths. Although, in principle, the incorporation of non-canonical amino acids into the core of proteins may affect their 3D structure and stability, our mutants of the protein PpiB maintained the wild-type structure. Furthermore, perturbations of the protein structure can be minimized by mutating solvent-exposed residues. In particular, mutation to flexible TFA-Lys residues allows exploring a greater conformational space, where a $^{19}F–^{19}F$ contact between two TFA-Lys residues can serve as a more broadly applicable indicator of spatial proximity. As the tools for incorporating non-canonical amino acids into proteins have become readily available, through-space $^{19}F–^{19}F$ couplings between CF_3 groups present a practical and sensitive tool for interrogating

specific interactions without the need for cumbersome resonance assignments of whole proteins including amino acid side chains.

EXPERIMENTAL SECTION

NMR spectra

All NMR spectra were recorded at 25 ºC on a Bruker 400 MHz NMR spectrometer equipped with a room temperature broadband probe designed for ^{19}F detection on the inner coil. Spectra of PpiB CF_3-Tyr(27,98) were recorded of 0.7 and 1 mM samples in NMR buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 90% H_2O/10% D_2O) with 0.5 mM trifluoroacetate as inner ^{19}F reference). Spectra of PpiB CF_3-Phe(27,98) and PpiB CF_3-Tyr(4,27,98) were recorded of 0.5 mM solutions in NMR buffer. The sample of PpiB F4A/CF_3-Phe(27,98) had a concentration of 0.6 mM in NMR buffer. Spectra of AncCDT-1 were recorded of 150 μM solutions in NMR buffer in a 3 mm NMR tube. Spectra of mRFP were recorded of 1.6 mM solution in phosphate buffered saline (PBS) in 5 mm NMR tube.

NMR simulations

NMR spectra were simulated using the program Spinach.[38] The calculations were performed for a spin system with two CF_3 groups with geometry taken from chain B of the crystal structure shown in Figure 5. Simulations assumed an axially symmetric CSA tensor with $\Delta = 25$ ppm and the main principal axis aligned with the O–C bond. To the best of our knowledge, a CSA tensor magnitude for CF_3-tyrosine has not been determined experimentally. Solid-state NMR measurements of 4-CF_3-phenylalanine have been reported, which yielded a CSA tensor with $\Delta = 39$ ppm.[56] Our simulations used $\Delta = 25$ ppm to improve the match in appearance of the DQF-COSY diagonal peaks with those observed in the experimental spectrum of Figure 2a (see Figure S11 for a series of DQF-COSY spectra simulated with different CSA tensor anisotropies and rotational correlation times). Redfield relaxation theory was applied in the laboratory frame with full tensor connectivity and a rotational correlation time of 9 ns.

ASSOCIATED CONTENT

Supporting Information

Experimental procedures for protein expression, purification, and crystallization; mass spectra of proteins produced; X-ray crystal structure statistics; selection of a new aminoacyl-tRNA synthetase for TFA-Lys; NMR data measuring ^{19}F–^{19}F correlations, $^{TS}J_{FF}$,

and $R_2(^{19}F)$ relaxation; NMR simulations for TOCSY build-up; simulations of $^{19}F–^{19}F$ DQF-COSY spectra, and NMR pulse program code.

The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

* Gottfried Otting - ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Email: gottfried.otting@anu.edu.au

Author Contributions

‡ H.W.O. and H.Q. contributed equally to this work.

ACKNOWLEDGMENT

We thank Prof. Peter G. Schultz for the pEVOL vector with the *p*-cyanophenylalanyl-tRNA synthetase. The research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. Financial support by the Australian Research Council for a Laureate Fellowship to G.O. (FL170100019), project funding (DP200100348 and DP21010088), and through a Centre of Excellence (CE200100012) is gratefully acknowledged.

ABBREVIATIONS

CSA = chemical shift anisotropy
GB1 = immunoglobulin-binding domain of protein G
NMR = nuclear magnetic resonance
NOE = nuclear Overhauser effect
PpiB = peptidyl-prolyl *cis/trans*-isomerase B
TFA-Lys = N^6-(trifluoromethyl)-L-lysine

REFERENCES

(1) Gerig, J. T. Fluorine NMR of proteins. *Prog. NMR Spectrosc*. **1994**, *26*, 293–370.
(2) Danielson, M. A.; Falke, J. J. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct*. **1996**, *25*, 163–195.
(3) Gakh, Y. G.; Gakh, A. A.; Gronenborn, A. M. Fluorine as an NMR probe for structural studies of chemical and biological systems. *Magn. Reson. Chem*. **2000**, *38*, 551–558.
(4) Kitevski-LeBlanc, J. L.; Prosser, R. S. Current applications of ^{19}F NMR to studies of protein structure and dynamics. *Prog. NMR Spectrosc*. **2012**, *62*, 1–33.
(5) Ye, L.; Larda, S. T.; Li, Y. F. F.; Manglik, A.; Prosser, R. S. A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in ^{19}F NMR studies of proteins. *J. Biomol. NMR* **2015**, *62*, 97–103.
(6) 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.* **2015**, *565*, 67–95.
(7) Arntson, K. E.; Pomerantz, W. C. K. Protein-observed fluorine NMR: a bioorthogonal approach for small molecule discovery. *J. Med. Chem*. **2016**, *59*, 5158–5171.
(8) Di Pietrantonio, C.; Pandey, A.; Gould, J.; Hasabnis, A.; Prosser, R. S. Understanding protein function through an ensemble description: characterization of functional states by ^{19}F NMR. *Methods Enzymol*. **2019**, *615*, 103–130.
(9) Divakaran, A.; Kirberger, S. E.; Pomerantz, W. C. K. SAR by (protein-observed) ^{19}F NMR. *Acc. Chem. Res*. **2019**, *52*, 3407–3418.
(10) Dalvit, C.; Vulpetti, A. Ligand-based fluorine NMR screening: principles and applications in drug discovery projects. *J. Med. Chem*. **2019**, *62*, 2218–2244.
(11) Mei, H.; Han, J.; Fustero, S.; Medio-Simon, M.; Sedgwick, D. M.; Santi, C.; Ruzziconi, R.; Soloshonok, V. A. Fluorine-containing drugs approved by the FDA in 2018. *Chem. Eur. J*. **2019**, *25*, 11797–11819.
(12) Kimber, B. J.; Feeney, J.; Roberts, G. C. K.; Birdsall, B.; Griffiths, D. V.; Burgen, A. S. V.; Sykes, B. D. Proximity of two tryptophan residues in dihydrofolate reductase determined by ^{19}F NMR. *Nature* **1978**, *271*, 184–185.
(13) Emsley, J. W.; Phillips, L.; Wray, V. Fluorine coupling constants. *Prog. NMR Spectrosc*. **1976**, *10*, 83–756.
(14) Haase, L.; Weisz, K. Switching the type of V-loop in sugar-modified G-quadruplexes through altered fluorine interactions. *Chem. Commun*. **2020**, *56*, 4539–4542.
(15) Arnold, W. D.; Mao, J.; Sun, H.; Oldfield, E. Computation of through-space $^{19}F–^{19}F$ scalar couplings via density functional theory. *J. Am. Chem. Soc*. **2000**, *122*, 12164–12168.
(16) Tuttle, T.; Gräfenstein, J.; Cremer, D. Analysis of the NMR through-space coupling mechanism between ^{19}F atoms. *Chem. Phys. Lett*. **2004**, *394*, 5–13.
(17) Hierso, J.-C. Indirect nonbonded nuclear spin−spin coupling: a guide for the recognition and understanding of "through-space" NMR *J* constants in small organic, organometallic, and coordination compounds. *Chem. Rev*. **2014**, *114*, 4838–4867.
(18) Perras, F. A.; Marion, D.; Boisbouvier, J.; Bryce, D. L.; Plevin, M. J. Observation of CH···π interactions between methyl and carbonyl groups in proteins. *Angew. Chem. Int. Ed*. **2017**, *56*, 7564–7567.
(19) Cordier, F.; Grzesiek, S. Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J. Am. Chem. Soc*. **1999**, *121*, 1601–1602.

(20) Cornilescu, G.; Hu, J.-S.; Bax, A. Identification of the hydrogen bonding network in a protein by scalar couplings. *J. Am. Chem. Soc*. **1999**, *121*, 2949–2950.
(21) Gemmecker, G. Direct detection of hydrogen bonds in biopolymers by NMR spectroscopy. *Angew. Chem. Int. Ed*. **2000**, *39*, 1224–1226.
(22) Arnold, W. D.; Oldfield, E. The chemical nature of hydrogen bonding in proteins via NMR: J-couplings, chemical shifts, and AIM theory. *J. Am. Chem. Soc*. **2000**, *122*, 12835–12841.
(23) Grzesiek, S.; Cordier, F.; Jaravine, V.; Barfield, M. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog. NMR Spectrosc*. **2004**, *45*, 275–300.
(24) Blake, P. R.; Lee, B.; Summers, M. F.; Adams, M. W. W.; Park, J.-B.; Zhou, Z. H.; Bax, A. Quantitative measurement of small through-hydrogen-bond and 'through-space' ^{1}H-^{113}Cd and ^{1}H-^{199}Hg J couplings in metal-substituted rubredoxin from *Pyrococcus furiosus*. *J. Biomol. NMR* **1992**, *2*, 527–533.
(25) Liu, J. J.; Horst, R.; Katrich, V.; Stevens, R. C.; Wüthrich, K. Biased signaling pathways in β_2-adrenergic receptor characterized by ^{19}F-NMR. *Science* **2012**, *335*, 1106–1110.
(26) Manglik, A.; Kim, T. H.; Masureel, M.; Altenbach, C.; Yang, Z.; Hilger, D.; Lerch, M. T.; Kobilka, T. S.; Thian, F. S.; Hubbell, W. L.; Prosser, R. S.; Kobilka, B. K. Structural insights into the dynamic process of b_2-adrenergic receptor signaling. *Cell* **2015**, *161*, 1101–1111.
(27) Wang, H. X.; Hu, W. H.; Liu, D. S.; Wüthrich, K. Design and preparation of the class B G protein-coupled receptors GLP-1R and GCFR for ^{19}F-NMR studies in solution. *FEBS J*. **2021**, *288*, 4053–4063.
(28) Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem*. **2010**, *79*, 413–444.
(29) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **2014**, *1844*, 1059–1070.
(30) Qianzhu, H.; Welegedara, A. P.; Williamson, H.; McGrath, A. E.; Mahawaththa, M. C.; Dixon, N. E.; Otting, G.; Huber, T. Genetic encoding of *para*-pentafluorosulfanyl phenylalanine: A highly hydrophobic and strongly electronegative group for stable protein interactions. *J. Am. Chem. Soc.* **2020**, *142*, 17277–17281.
(31) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc*. **2007**, *129*, 1160–1166.
(32) Li, C.; Wang, G.; Wang, Y.; Creager-Allen, R.; Lutz, E. A.; Scronce, H.; Slade, K. M.; Ruf, R. A. S.; Mehl, R. A.; Pielak, G. J. Protein ^{19}F NMR in *Escherichia coli*. *J. Am. Chem. Soc*. **2010**, *132*, 321–327.
(33) Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H. *In vivo* incorporation of non-canonical amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc*. **2008**, *130*, 9268–9281.
(34) Loscha, K. V.; Herlt, A. J.; Qi, R.; Huber, T.; Ozawa, K.; Otting, G. Multiple-site labeling of proteins with non-canonical amino acids. *Angew. Chem. Int. Ed.* **2012**, *51*, 2243–2246.
(35) Clifton, B. E. & Jackson, C. J. Ancestral protein reconstruction yields insight into adaptive evolution of binding specificity in solute-binding protein. *Cell Chem. Biol*. **2016**, *23*, 236–245.
(36) Clark, K. J.; Balciunas, D.; Pogoda, H.-M.; Ding, Y.; Westcot, S. E.; Bedell, V. M.; Greenwood, T. M.; Urban, M. D.; Skuster, K. J.; Petzold, A. M.; Ni, J.; Nielsen, A. L.; Patowary, A.; Scaria, V.; Sivasubbu, S.; Xu, X.; Hammerschmidt, M.; Ekker, S. C. *In vivo* protein trapping produces functional expression codex of the vertebrate proteome. *Nat. Methods*, **2011**, *8*, 506–512.
(37) Edwards, K. J.; Ollis, D. L.; Dixon, N. E. Crystal structure of cytoplasmic *Escherichia coli* peptidyl-prolyl isomerase: evidence for decreased mobility of loops upon complexation. *J. Mol. Biol*. **1997**, *271*, 258–265.
(38) Hogben, H. J.; Krzystyniak, M.; Charnock, G. T. P.; Hore, P. J.; Kuprov, I. Spinach – A software library for simulation of spin dynamics in large spin systems. *J. Magn. Reson*. **2011**, *208*, 179–194.
(39) Yu, L.; Hajduk, P. J.; Mack, J.; Olejniczak, E. T. Structural studies of Bcl-xL/ligand complexes using ^{19}F NMR. *J. Biomol. NMR* **2006**, *34*, 221–227.
(40) Zwahlen, C.; Vincent, S. J. F.; Di Bari, L.; Levitt, M. H.; Bodenhausen, G. Quenching spin diffusion in selective measurements of transient Overhauser effects in nuclear magnetic resonance. Applications to oligonucleotides. *J. Am. Chem. Soc*. **1994**, *116*, 362–368.
(41) Neuhaus, D.; Williamson, M. P. The Nuclear Overhauser Effect in Structural and Conformational Analysis, Wiley-VCH, 2000.
(42) Plateau, P.; Guéron, M. Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc*. **1982**, *104*, 7310–7311.
(43) Zhang, F.; Zhou, Q.; Yang, G.; An, L.; Li, F.; Wang, J. A genetically encoded ^{19}F NMR probe for lysine acetylation. *Chem. Commun*. **2018**, *54*, 3879–3882.
(44) Abdelkader, E. H.; Qianzhu, H.; George, J.; Frkic, R.; Jackson, C.; Nitsche, C.; Otting, G.; Huber, T. Genetic encoding of cyanopyridylalanine for in-cell protein macrocyclization by the nitrile-aminothiol click reaction. ChemRxiv, Preprint, 10-27-**2021**. DOI: 10.33774/chemrxiv-2021-r5dn6.
(45) Qianzhu, H.; Welegedara, A. P.; Williamson, H.; McGrath A. E.; Mahawaththa, M. C.; Dixon, N. E.; Otting, G.; Huber, T. Genetic encoding of *para*-pentafluorsulfanyl phenylalanine: a highly hydrophobic and strongly electronegative group for stable protein interactions. *J. Am. Chem. Soc*. **2020**, *142*, 17277−17281.
(46) Shu, X.; Wang, L.; Colip, L.; Kallio, K.; Remington, S. J. Unique interactions between the chromophore and glutamate 16 lead to far-red emission in a red fluorescent protein. *Protein Sci.* **2009**, *18*, 460−466.
(47) Ernst, L.; Ibrom, K. A new quantitative description of the distance dependence of through-space ^{19}F, ^{19}F spin–spin coupling. *Angew. Chem. Int. Ed*. **1995**, *34*, 1881–1882.
(48) Mallory, F. B.; Mallory, C. W.; Butler, K. E.; Lewis, M. B.; Xia, A. Q.; Luzik, E. D., Jr.; Fredenburgh, L. E.; Ramanjulu, M. M.; Van, Q. N.; Francl, M. M.; Freed, D. A.; Wray, C. C.; Hann, C.; Nerz-Stormes, M.; Carroll, P. J.; Chirlian, L. E. Nuclear spin–spin coupling via nonbonded interactions. 8. The distance dependence of through-space fluorine–fluorine coupling. *J. Am. Chem. Soc*. **2000**, *122*, 4108–4116.
(49) Mallory, F. B.; Mallory C. W.; Baker M. B. Nuclear spin–spin coupling via nonbonded interactions. 6. F–F coupling through an intervening phenyl group. *J. Am. Chem. Soc.* **1990**, *112*, 2577–258.
(49) 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*, 583–589.
(50) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Rie Lee, G.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Erbrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.
(51) Otting, G.; Liepinsh, E.; Wüthrich, K. Protein hydration in aqueous solution. *Science* **1991**, *254*, 974–980.
(52) Abdelkader, E. H.; Qianzhu, H.;Tan, Y. J.; Adams, L. A.; Huber, T.; Otting, G. Genetic encoding of N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine for NMR studies of protein–protein and protein–ligand interactions. *J. Am. Chem. Soc.* **2021**, *143*, 1133–1143.
(53) Salwiczek, M.; Nyakatura, E. K.; Gerling, I. M.; Ye, S.; Koksch, B. Fluorinated amino acids: compatibility with native protein structures and effects on protein–protein interactions. Chem. Soc. Rev. 2012, 41, 2135–2171.
(54) Didenko, T.; Liu, J. J.; Horst, R.; Stevens, R. C.; Wüthrich, K. Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Curr. Opin. Struct. Biol.* **2013**, *23*, 740–747.
(55) Picard, L.-P.; Prosser R. S. Advances in the study of GPCRs by ^{19}F NMR. *Curr. Opin. Struct. Biol.* **2021**, *69*, 169–176.
(56) Grage, S.L.; Dürr, U. H. N.; Afonin, S.; Mikhailiuk, P. K.; Komarov, I. V.; Ulrich, A. S. Solid state ^{19}F NMR parameters of fluorine-labeled amino acids. Part II: Aliphatic substituents. *J. Magn. Reson*. **2008**, *191*, 16–23.