1.3 Å Crystal Structure of E. coli Peptidyl–Prolyl Isomerase B with Uniform Substitution of Valine by (2 S ,3 S)-4-Fluorovaline Reveals Structure Conservation and Multiple Staggered Rotamers of CH $_2$ F Groups

DOI:
10.1021/acs.biochem.4c00345

Document Version
Accepted author manuscript

Published in:
Biochemistry

1.3 Å Crystal Structure of *E. coli* Peptidyl–Prolyl Isomerase B with Uniform Substitution of Valine by (2*S*,3*S*)-4-Fluorovaline Reveals Structure Conservation and Multiple Staggered Rotamers of CH_2F Groups

Rebecca L. Frkic,[†] Yi Jiun Tan,[†] Ansis Maleckis,[‡] Nicholas F. Chilton,[§] Gottfried Otting,[†,*] Colin J. Jackson[†,*]

* Email: colin.jackson@anu.edu.au and gottfried.otting@anu.edu.au

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

[‡] Latvian Institute of Organic Synthesis, Aizkraukles 21, LV-1006 Riga, Latvia

[§] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia; Department of Chemistry, The University of Manchester, Manchester, M13 9PL, United Kingdom

Keywords: 4-fluorovaline; fluorine; ^{19}F-NMR spectroscopy; *E. coli* peptidyl–prolyl *cis*–*trans* isomerase B (PpiB); nuclear magnetic resonance; X-ray crystallography

Abstract

(2*S*,3*S*)-4-Fluorovaline (FVal) is an analogue of valine, where a single CH_3 group is substituted by a CH_2F group. In the absence of valine, *E. coli* valyl-tRNA synthetase uses FVal as a substitute, enabling the production of proteins uniformly labeled with FVal. Here we describe the production and analysis of *E. coli* peptidyl–prolyl isomerase B where all sixteen valine residues have been replaced by FVal synthesized with a ^{13}C-labeled CH_2F group. Although the melting temperature is lower by about 11 °C relative to the wild-type protein, the three-dimensional protein structure is almost completely conserved as shown by X-ray crystallography. The CH_2F groups invariably populate staggered rotamers. Most CH_2F groups populate two different rotamers. The increased space requirement of fluorine versus hydrogen does not prohibit rotamers that position fluorine next to a backbone carbonyl carbon. ^{19}F NMR spectra show a signal dispersion over 25 ppm. The most high-field shifted ^{19}F resonances correlate with large $^3J_{HF}$ coupling constants, confirming the impact of the γ-gauche effect on the signal dispersion. The present work is the second

experimental verification of the effect and extends its validity to fluorovaline. The abundance of valine in proteins and structural conservation with FVal renders this valine analogue attractive for probing proteins by ^{19}F NMR spectroscopy.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy of proteins containing site-specifically installed fluorine labels provides outstanding opportunities for site-specific interrogation of protein structure, function, and dynamics. ^{19}F spins yield highly sensitive NMR spectra because ^{19}F is the only naturally occurring isotope, possesses a spin ½ and a high Larmor frequency, features an unusually large chemical shift dispersion and, owing to the absence of fluorine from most biological macromolecules, produces background-free ^{19}F-NMR spectra of target proteins labeled with fluorine. Fluorine atoms can readily be introduced by producing proteins from amino acid mixtures, where a single amino acid is included in fluorinated form. Proteins produced with global substitution of aromatic amino acids with fluorinated analogues have long been used to monitor protein folding and interactions with ligands by ^{19}F NMR.[1,2]

Global substitution of amino acids by their fluorinated analogues leads to minimal structural perturbation as the van der Waals radius of fluorine is not much larger than that of hydrogen (1.47 Å versus 1.2 Å), and a C–F bond is only little longer than a C–H bond (1.35 Å versus 1.09 Å). In addition, carbon-bonded fluorine readily engages in hydrophobic interactions and does not form strong hydrogen bonds,[3-6] so that CH_2F groups can readily be accommodated in the hydrophobic core of proteins despite their polarity.[7-9]

As the side chains of aromatic amino acids are relatively large and rigid, they have few options to adjust their conformation to accommodate the greater steric demands of a fluorine versus hydrogen atom. Lesser structural impact would be expected for amino acids with only a singly fluorinated methyl group, as the CH_2F group can avoid steric constraints by adopting any of three different staggered rotamers. We recently confirmed that CH_2F groups indeed populate staggered rotamers in high-resolution crystal structures of the *E. coli* protein peptidyl–prolyl *cis–trans* isomerase B (PpiB) produced with different isomers of fluoroleucine, comprising (2*S*,4*S*)-5-fluoroleucine, (2*S*,4*R*)-fluoroleucine, and 5,5'-difluoroleucine.[8] The crystal structures revealed near complete conservation of the backbone and amino acid side chain conformations. In addition, different protein chains in

the unit cells of the crystals frequently showed the same CH_2F group populating different rotamers, indicating facile rotation. In agreement with this result, ^{19}F NMR data indicated rapid exchange between different rotamers in solution.[7]

In the present work, we explore the structural impact of (2*S*,3*S*)-4-fluorovaline (FVal) residues in PpiB. The isopropyl group of the side chain of valine is closer to the backbone atoms than in leucine, potentially leaving fewer options to satisfy the steric demands and polarity of the C–F group by conformational adjustments of the side chain. Valine and glycine are the most abundant amino acids in PpiB (16 residues each compared to 5 leucine residues), presenting a challenging example for the structural conservation of a protein with multiple aliphatic CH_2F groups. Finally, we use ^{19}F NMR spectra to assess whether the γ-gauche effect applies to FVal. The γ-gauche effect predicts that the ^{19}F chemical shift of a CH_2F group is high-field for rotamers positioning the fluorine atom *trans* relative to a hydrogen atom and low-field if it is *trans* relative to a carbon atom.[10] In the case of FVal, it predicts that the $^3J_{HF}$ coupling with the H^β atom of the isopropyl group inversely correlates with the ^{19}F chemical shift.[7]

Materials/Experimental Details

Protein production and purification. FVal was synthesized with a ^{13}C labeled CH_2F group as described previously.[9] The construct of PpiB was the same as used previously for crystallization of the protein with fluorinated leucines and contained a C-terminal His_6 tag.[8] The protein was produced by continuous exchange cell-free protein synthesis following the published protocol,[7] except that valine was omitted from the amino acid mixture and FVal was provided instead at 1 mM concentration. Cell-free protein synthesis (CFPS) was conducted in two parallel reactions, each with 1 mL inner reaction mixture and 10 mL outer buffer. The protein purification followed the published protocol.[7]

Protein crystallization. The protein was concentrated to 8 mg/mL in 50 mM imidazole pH 7.0, 1 mM EDTA, and 1 mM DTT using a 10 kDa MWCO centrifugal concentrator prior to crystallization. Crystals were obtained after approximately one week using the hanging drop vapor diffusion method against a reservoir solution containing 31% PEG 3350, 0.1 M Tris-HCl pH 7.0, and 200 mM sodium acetate trihydrate. The crystallization drop was comprised of 1.5 µL protein and 1 µL well solution. Crystals were flash cooled

in liquid nitrogen without further cryoprotecting. A table of crystallographic data statistics can be found in Table S1, and the coordinates were deposited to the Protein Data Bank under accession ID 9C5D.

Structure determination. Data were collected on the MX2 beamline at the Australian Synchrotron.[11] Diffraction images were processed using XDS[12] in space group P 2_1 2_1 2_1. The data were truncated using Aimless[13] until statistics were of suitable quality in the outer shell. The structures were solved by molecular replacement (Phaser MR, CCP4)[14] using the structure of wild-type PpiB (PDB ID 2NUL)[15] as the search model. Non-canonical amino acids were built using PyMOL.[16] Geometry restraints were generated using Phenix.eLBOW.[17] Refinement was performed using Phenix.refine.[18] The program Coot[19] was used for multiple rounds of refinement and model building until *R*-factors converged and the electron density was modeled as best as possible. MolProbity[20] was used for structure validation. Figures were created using PyMOL 2.0.

NMR spectroscopy. All NMR spectra were recorded at 25 ºC on a Bruker 500 MHz NMR spectrometer equipped with a $^{1}H/^{19}F/^{13}C$ cryoprobe. The sample concentration was 0.6 mM in NMR buffer (50 mM HEPES, pH 7.5, 100 mM NaCl) containing 10% D_2O. 1D ^{19}F NMR spectra were recorded using an acquisition time of 123 ms, a recovery delay of 2 s, 256 scans, and ^{1}H decoupling using Waltz16. The [^{13}C,^{19}F]-HSQC spectrum was recorded using INEPT and reverse INEPT delays of 3 ms, t_{1max} = 14 ms, t_{2max} = 74 ms, a 1 s recovery delay, and a total experimental time of 1.5 h. Waltz16 decoupling was applied to ^{1}H during the t_1 and t_2 acquisition times and garp decoupling was applied to ^{13}C during t_2. The spectrum was processed with exponential window multiplication (30 Hz line broadening) in the ^{19}F dimension and a cosine squared window after doubling the data by linear prediction in the ^{13}C dimension. $^{3}J_{HF}$ coupling constants were measured using the previously published short-delay ^{1}H,^{19}F-correlation experiment as described,[7] using Δ_1 = 7 ms, Δ_2 = 2.5 ms, t_{1max} = 7 ms, t_{2max} = 74 ms, a 1 s recovery delay, Waltz16 ^{13}C decoupling during the evolution and detection times, and a total experimental time of 8 h. The spectrum was processed with exponential window multiplication (50 Hz line broadening) in the ^{19}F dimension and a cosine squared window after doubling the data by linear prediction in the ^{1}H dimension.

DFT calculations of ^{19}F chemical shifts and $^{3}J_{HF}$ coupling constants

Restricted Kohn-Sham density-functional theory (DFT) calculations were performed using the ωB97X-D3BJ functional,[21] D3BJ dispersion corrections,[22,23] the aug-cc-pvdz basis set,[24] and the RIJCOSX approximation for two-electron integrals[25–27] with on-the-fly auxiliary basis sets,[28] in Orca 6.0.0.[29,30] Each valine residue was taken from the crystal structure of wild-type PpiB (PDB ID: 2NUL)[15] with the peptide backbone fixed and truncated as $NH_2COCH(R)NHCOH$, where R is the $CH(CH_2F)(CH_3)$ group of FVal which was optimized along with the truncating backbone H atoms. The geometry optimization was performed in vacuum, omitting all other amino acids and solvent, however in the presence of a solvent continuum[30] with a dielectric constant of 6.5 and a refractive index of 1.4 to mimic the interior of a protein. Geometry optimizations were performed for three staggered rotamers of the CH_2F group of FVal for each residue, giving the optimized χ_2 angles and energies. Calculations for the ^{19}F chemical shift and $^{3}J_{HF}$ coupling constants for each rotamer of each residue were performed using the gauge-including atomic orbital (GIAO) method.[31]

Results

Protein expression and purification. FVal was synthesized with a ^{13}C label in the CH_2F group as described previously.[9] The total amount of FVal available was 7 mg. To minimize the availability of canonical valine and avoid any toxicity of FVal on bacterial growth, we produced PpiB by cell-free protein synthesis (CFPS) using a dialysis system with a total of 2 mL inner reaction mixture and 20 mL outer reaction mixture as described previously,[7,32] except that valine was omitted from the amino acid mixture and FVal was provided at a concentration of 1 mM. The CFPS reaction yielded 6 mg of purified PpiB. Intact protein mass spectrometry indicated that the most abundant protein species contained 16 FVal residues, but the presence of species with 15 and 14 FVal residues indicates that each of the 16 valine sites in the amino acid sequence also featured valine with an average likelihood of about 3.5% (Figure S1). In the following, we refer to the PpiB construct produced with global substitution of valine with (2*S*,3*S*)-4-fluorovaline as PpiB-FVal.

Protein stability. The stability of PpiB-FVal against thermal denaturation was probed by a commercial protein thermal shift assay (PTS; Figure S2). The melting temperature T_m determined from fitting a two-state model (folded and unfolded states) was 41 °C, whereas the T_m value of the wild-type protein measured in this way was 52 °C. Although the PTS curves suggest that the proteins are stable at 25 °C, some PpiB-FVal precipitated during the NMR experiments, whereas solutions of the wild-type protein are stable under the same conditions. This suggests that the kinetic stability of PpiB-FVal at 25 °C is also lower than for the wild-type protein.

X-ray crystal structure. The crystal structure of PpiB-FVal was determined with a resolution of 1.3 Å. It contains four protein chains per asymmetric unit. The backbone atoms of all four chains superimpose closely with the crystal structure of the wild-type protein (PDB ID 2NUL)[15] with a root-mean-square deviation (RMSD) below 0.4 Å (Figure 1). With a single exception, the amino acid side chains of the buried residues other than FVal are fully conserved and assume the same conformations as in the wild-type protein. The exception is the side chain of Phe4, which is isomorphous in chains A, B, and D, but not in chain C, where the dihedral angle χ_2 is changed by almost 90° (Figure S3). The aromatic ring is next to the side chain of FVal160, which shows two different χ_1 angles in different chains, one of which is the same as in the wild-type protein. Two different χ_1 angles are also observed for the side chain of FVal2, which is located near FVal160. In all other cases, the carbon atoms of the FVal side chains assume the same conformations in PpiB-FVal as the corresponding valine residues in wild-type PpiB. Notably, despite the non-uniform distribution of FVal sites across the protein (Figures 1 and S4), the structure allows direct fluorine–fluorine contacts only for residues 2, 160, and 162, and for the pair of FVal residues in positions 102 and 133.

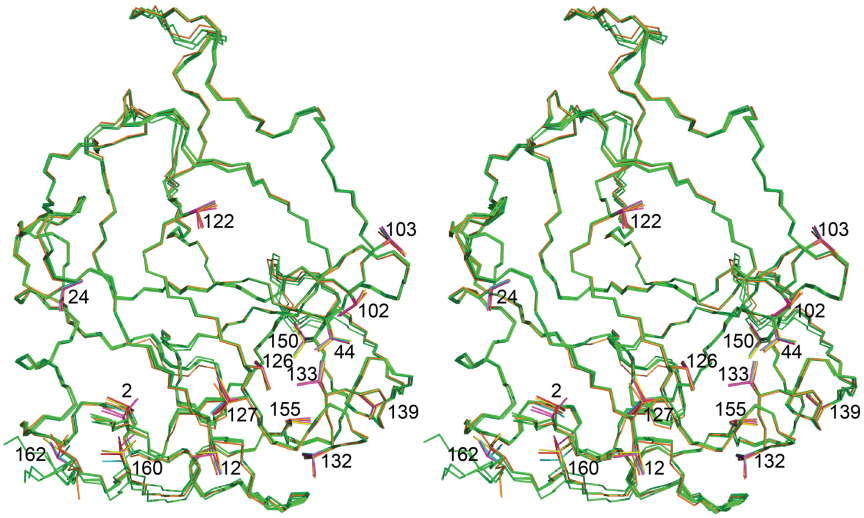

Figure 1. Stereoview of a superimposition of the four identical chains A–D in the unit cell of the crystal structure of PpiB-FVal with wild-type PpiB (PDB ID 2NUL[15]), demonstrating high structural conservation. The backbone and Val side chains of the wild-type protein are shown in orange. The structure of PpiB-FVal is shown with the backbone of the chains A–D in green and the carbon atoms of the FVal side chains in red, cyan, magenta, and yellow, respectively, with their sequence numbers identified.

The positions of the fluorine atoms were determined by modeling into clear mF_O-DF_C difference density. Figures 2 and S5 show that the C–F bonds of the FVal residues assume staggered conformations in all cases where the position of the fluorine atom could be confidently established by good electron density. Most of the FVal residues displaying a well-defined conformation in chain A of the crystal structure feature at least one alternative staggered rotamer for the CH_2F group in a different chain (Table S2). Importantly, the structure refinement protocol used did not restrain the C–F bonds to staggered conformations by any explicit geometry restraints, enabling the FVal residues to adopt the dihedral angles that best fit the data.

Given the conservation of the protein structure and the limited rotational freedom of the FVal side chains, all fluorine atoms are out of reach of an H-bond donor group of another amino acid sidechain. While this does not exclude H-bonds with water molecules for FVal residues on the surface of the protein, there is no evidence that the CH_2F rotamers populated seek to either expose the fluorine to solvent or protect it from solvent access. This confirms the weak H-bonding capacity of the C–F bonds.[3-6]

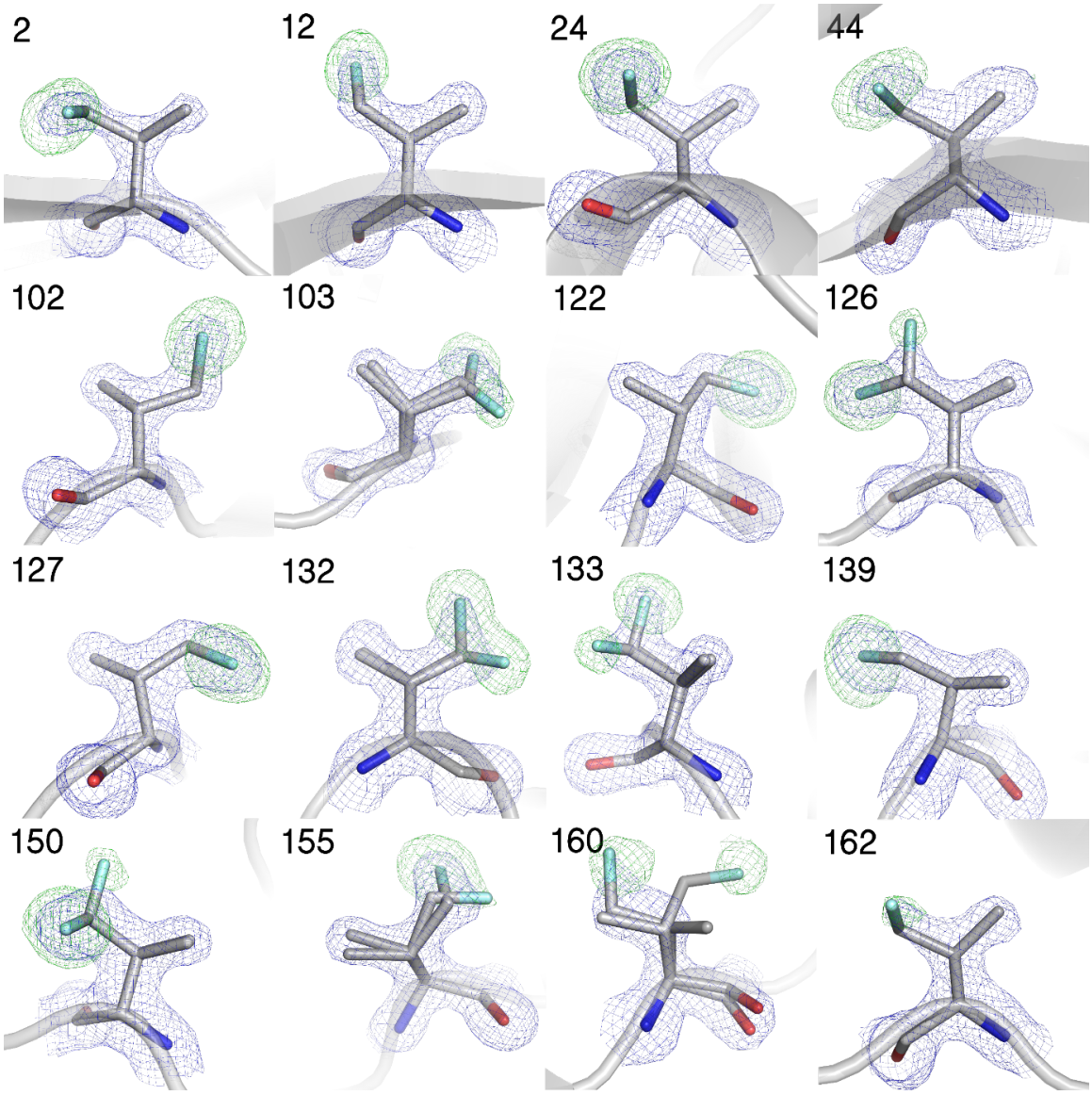

Figure 2. Electron density of fluorinated valine residues (grey sticks) observed for chain A. Omit density (modeled without fluorine atoms) is shown as green mesh (contoured to 4σ), with blue mesh showing 2mF_O-DF_C density (contoured to 2σ). The electron densities demonstrate that the CH_2F groups preferentially populate one or two staggered rotamers.

^{19}F-NMR. The 1D ^{19}F NMR spectrum displays signals dispersed over more than 20 ppm, demonstrating exceptional sensitivity of the ^{19}F NMR chemical shifts to the local chemical environment (Figure 3). There are more than 16 peaks and their integrals differ. This sample heterogeneity was also apparent in a [^{13}C,^{19}F]-HSQC spectrum that revealed cross-

peaks at more than 16 different ^{19}F chemical shifts (Figure 4). The heterogeneity in the ^{19}F NMR spectrum is more pronounced than expected from the low level of natural valine at each location of FVal, suggesting the presence of different conformations that exchange with each other at a rate that is slow on the chemical shift time scale (milliseconds). Apart from an intense signal at -222.5 ppm characteristic of a highly flexible solvent-exposed FVal residue, the natural line widths range from about 30 Hz (signal at -214.4 ppm) to over 150 Hz (signal at -228.3 ppm).

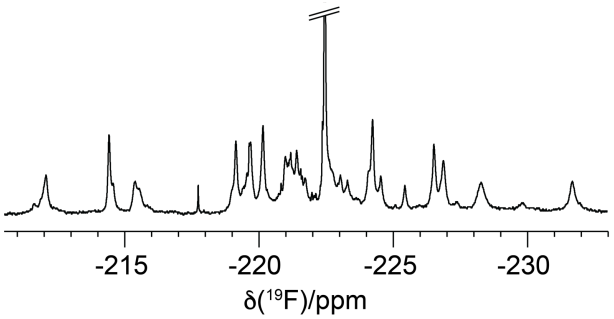

Figure 3. 1D ^{19}F NMR spectrum of *E. coli* PpiB with valine globally substituted by (2*S*,3*S*)-4-fluorovaline (PpiB-FVal). Spectrum recorded with ^{1}H and ^{13}C decoupling at 25 ^{o}C on a 500 MHz NMR spectrometer and processed with 10 Hz line broadening. The tall and narrow peak in the center of the spectrum appears to stem from degraded protein as it grew in intensity during prolonged NMR measurements, which were also accompanied by protein precipitation.

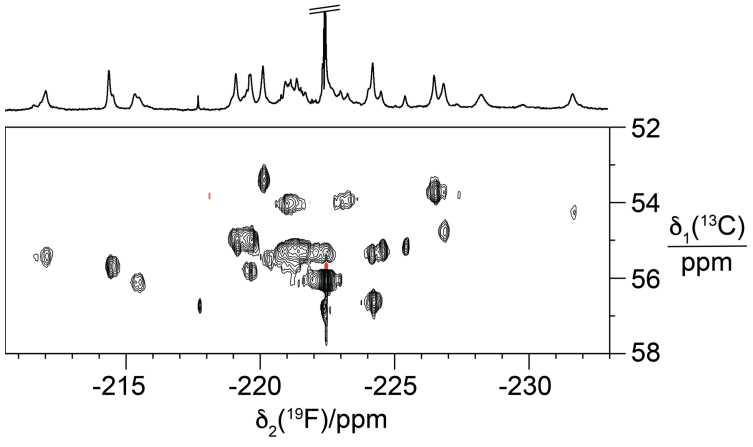

Figure 4. [^{13}C,^{19}F]-HSQC spectrum of PpiB-FVal. The spectrum was recorded at 25 ^{o}C on a Bruker 500 MHz NMR spectrometer with ^{1}H decoupling during the evolution and

detection times. The 1D ^{19}F NMR spectrum recorded with ^{1}H and ^{13}C decoupling is shown at the top.

Despite the broad line widths in the ^{19}F NMR spectrum, a short-delay $^{1}H,^{19}F$-correlation experiment yielded cross-peaks for the fluorine atoms with the protons of the $C^{\gamma}H_2F$ and $C^{\beta}H$ groups (Figure 5). Cross-peaks in this experiment arise from magnetization transfer between ^{1}H and ^{19}F via $^{2}J_{HF}$ and $^{3}J_{HF}$ couplings. Denoting the integrals of the cross-peaks with the protons of the $C^{\beta}H$ (1–3 ppm) and $C^{\gamma}H_2F$ (3.5–5 ppm) groups as I_{3bond} and I_{2bond}, respectively, the ratio I_{3bond}/I_{2bond} can be used to estimate the size of the $^{3}J_{HF}$ coupling constant.[7] The $^{3}J_{HF}$ couplings estimated in this way for the most low-field and most high-field ^{19}F NMR signals are 19 and 48 Hz, respectively. These values are larger than predicted from the Karplus curve of $^{3}J_{HF}$ couplings for *gauche* (ca. 8 Hz) and *trans* (ca. 44 Hz) conformations.[33,34] The experiment would overestimate the $^{3}J_{HF}$ couplings, if the T_2 relaxation time were shorter for the H^{γ} than H^{β} protons. This may have been the case, as the FVal amino acid provided had been synthesized with a selective ^{13}C label in the CH_2F group but not the $C^{\beta}H$ group,[9] providing an additional source of dipolar relaxation for the protons of the CH_2F group. Most notably, the H^{β}–^{19}F cross-peaks with the greatest intensity and, hence, the largest $^{3}J_{HF}$ values are consistently found at the high-field end of the ^{19}F dimension. This correlation between ^{19}F chemical shift and $^{3}J_{HF}$ coupling constant is a hallmark of the γ-gauche effect.[7]

As the crystal structure demonstrates the prevalence of multiple rotamers for the CH_2F groups of most FVal residues, the chemical shift of ^{19}F NMR signals downfield from -225 ppm may reflect either a single conformation or an average due to rapid exchange between different rotamers.

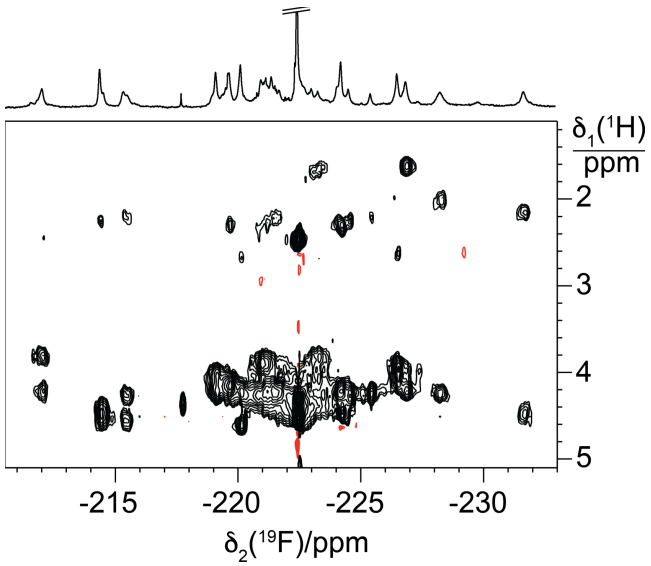

Figure 5. Short-delay $^{1}H,^{19}F$-correlation experiment of PpiB-FVal. The cross-peaks arise from $^{2}J_{HF}$ couplings within the $C^{\gamma}H_2F$ groups and from $^{3}J_{HF}$ couplings with the $C^{\beta}H$ groups of fluorovaline residues. The 1D ^{19}F NMR spectrum is plotted at the top.

Computed rotamer energies, ^{19}F shielding, and $^{3}J_{HF}$ coupling constants

The crystal structure of PpiB-FVal indicates that the CH_2F groups prefer staggered rotamer conformations (Figure S6). For 1-fluoropropane, no energy difference was reported between any of the three staggered conformations and the energy barriers between them were calculated to range between 4 and 5 kcal/mol.[10] In the case of PpiB-FVal, the proximity of the CH_2F groups to the protein backbone and concomitant steric constraints are expected to render the energies not only dependent on the χ_2 dihedral angle, but also on the χ_1 and backbone dihedral angles. To explore the significance of these effects, we performed DFT calculations, considering each of the sixteen FVal residues in isolation. Starting from the crystal structure of wild-type PpiB, the coordinates of the valine residues were extracted, hydrogen atoms were attached, and a fluorine atom was attached to the $C^{\gamma 1}$ atom. The C–F bond was oriented to create the three staggered rotamers with $\chi_2 \sim -60^{o}$, 60^{o}, and 180^{o}. Energy minimization of the amino acid side chain changed the χ_2 angles by 5 ± 7^{o} from the starting points. The conformation with the lowest energy was more likely to feature a χ_2 angle deviating from the starting point when the distance between the fluorine atom and the backbone atoms was short. The calculations indicated that the lowest energy state can feature any of the three rotamers of the CH_2F group, with energies differing by

less than 2 kcal/mol and usually less than 1 kcal/mol. In agreement with the γ-gauche effect,[10] the calculations predicted the most high-field shifted ^{19}F chemical shifts (by 15±7 ppm) for the rotamer positioning the C^{γ}–F bond *trans* relative to the C^{β}–H^{β} bond compared with the two *gauche* rotamers. The corresponding $^{3}J_{HF}$ coupling constants were calculated to be 30±1 Hz and 9±3 Hz, respectively (Table S3).

For the χ_2 = -60° rotamer, the ^{19}F chemical shift is particularly sensitive to the precise χ_2 angle, with a change from -60° to -70° calculated to increase the ^{19}F chemical shift by over 10 ppm (Table S3). In agreement with the DFT calculations, the FVal residues populating χ_2 = -60° rotamers in the crystal structure of PpiB-FVal show a trend for χ_2 towards more negative χ_2 angles (Figure S6). The calculations also predict that a χ_2 value of -75° is accompanied by $^{3}J_{HF}$ > 10 Hz. Therefore, the proximity of the backbone to the CH_2F group in FVal and the associated bias in the χ_2 angle away from -60° is expected to obscure the relationship between ^{19}F chemical shift and $^{3}J_{HF}$ coupling constant for all but the most high-field ^{19}F NMR signals.

Discussion

The successful production of protein with very high FVal enrichment was key to the ^{19}F NMR experiments of the present work, as the ^{19}F chemical shifts are sensitive to small differences in the chemical environment. Alanine and leucine containing fluorinated methyl groups are known to be toxic for bacterial cell growth.[10,35] FVal is likely to be equally toxic, making it difficult to grow an auxotrophic strain in minimal medium containing FVal in place of valine. In contrast, cell-free protein synthesis enabled obtaining protein with high FVal enrichment by simple substitution of valine for FVal without any special optimization of the system.

The present work shows that the structure of a globular protein is little affected by the global substitution of sixteen valine residues by a fluorinated analogue, although the kinetic and thermodynamic stability of the protein towards unfolding is reduced. Remarkably, also the amino acid side chains of all nearby non-valine residues retain their native conformation, with a ring rotation of Phe4 in one of the four molecules in the asymmetric unit being the single exception among the non-solvent exposed side chains. The high structural conservation extends to the carbon atoms of the FVal residues. Most of

the CH_2F groups, however, populate different rotamers, indicating that the rotation common to methyl groups is still quite possible for CH_2F groups. Nonetheless, the structure also features examples, where solvent-exposed CH_2F groups populate single rotamers even when there is no apparent steric hindrance to rotation. There is thus no clear correlation between the solvent exposure of a CH_2F group and the number of rotamers it populates in the crystal structure.

While the crystal structure was determined at cryogenic temperatures favoring low energy conformations, increased temperatures are expected to increase the populations of alternative rotamers of the CH_2F groups. In the case of PpiB produced with 5-fluoroleucines, average $^3J_{HF}$ coupling constants indicate rapid exchange of the CH_2F groups between *trans* and *gauche* rotamers.[7] In the case of PpiB-FVal, the rotation of the CH_2F groups may be more hindered due to closer proximity of the fluorine atom to the amino acid backbone. If the exchange between different rotamers is slow, this may explain why the 1D ^{19}F NMR spectrum of PpiB-FVal shows more than sixteen resonances. Preferential population of the *trans* rotamer is indicated for the most high-field shifted ^{19}F NMR signals as they feature the largest $^3J_{HF}$ coupling constants. This correlation is a hallmark of the γ-gauche effect, which predicts a large difference in ^{19}F chemical shifts between *trans* and *gauche* conformations.[10] Our results with PpiB-FVal confirm this effect, which has only recently been verified experimentally for the CH_2F rotamers of 5-fluoroleucine residues in PpiB.[7,8]

In general, the present analysis of FVal conformations in PpiB confirms the conclusions drawn from the previous analysis of two different diastereomers of 5-fluoroleucine (FLeu) and 5,5'-difluoroleucine in PpiB.[3] (i) The CH_2F groups assume staggered rotamers. (ii) Alternative staggered rotamers are readily populated. (iii) The rotamers show no obvious preference for the fluorine to be either shielded from or exposed to the solvent. (iv) The conformations of the carbon skeletons of the fluorinated amino acid side chains mostly reproduce those of the canonical residues in the wild-type protein. (v) There is no evidence for the fluorine atoms to engage with H-bond donating groups of other amino acid side chains.

As the side chain of valine is shorter than the side chain of leucine, there are side chain conformations of FVal that position the fluorine atom very close to the protein backbone. An intriguing situation arises for FVal residues that feature a dihedral angle χ_1

of 180°. As shown in Figure 6, this dihedral angle can lead to an intra-residual distance between the fluorine atom and the carbonyl carbon that is short compared with the commonly assumed van der Waals radii of fluorine (1.47 Å) and carbon (1.70 Å). Twelve of the sixteen FVal residues feature χ_1 = 180°. In seven of them, the CH_2F group populates the rotamer that positions the fluorine atom next to the carbonyl carbon. The geometry could be favored by a partial negative charge on the fluorine atom and a partial positive charge on the carbonyl carbon. The preference for this interaction does not seem to be strong, however, as this side chain conformation is always accompanied by other rotamers of the CH_2F group that do not position the fluorine near the carbonyl group. In the case of FVal122, DFT calculations suggest very similar energies for the χ_2 = -60° and 180° rotamers, indicating little energy penalty for the conformation shown in Figure 6.

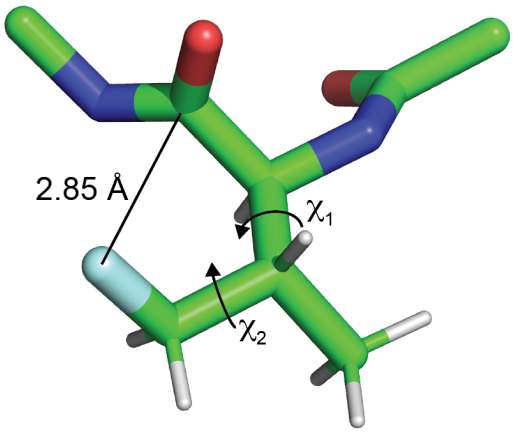

Figure 6. Main conformation of FVal122, showing that FVal residues can accommodate the fluorine atom next to their carbonyl carbon. The stick representation shows the bonds with heavy atoms and hydrogens as thick and thin lines, respectively. The intra-residual distance between the fluorine and carbonyl carbon atom is indicated. The display includes the carbonyl group and α-carbon of the preceding residue and the amino group and α-carbon of the following residue. The conformation shown corresponds to the side chain rotamers χ_1 = 180° and χ_2 = -60°.

High conservation of the three-dimensional protein structure following the installation of multiple FLeu or FVal residues is supported by the apparent ease with which most CH_2F groups populate different rotamers, generating different options to fit the fluorine atom into energetically optimal positions. Nonetheless, the presence of fluorine

can sometimes alter the amino acid side chain conformations in more fundamental ways, as evidenced by different χ_1 rotamers of the FVal residues in positions 2 and 160. The FVal residues in positions 2, 160, and 162 are sufficiently close for direct fluorine–fluorine contacts, which may be disfavored by the electric dipole moments of the CF groups. Conceivably, the chances of structural perturbation and destabilization towards protein unfolding by the presence of multiple fluorinated amino acids can be avoided by producing protein with only a low level of valine to FVal substitution, but the sensitivity of the ^{19}F NMR experiments would suffer.[36]

In conclusion, the substitution of a CH group by a CF group in one of the methyl groups of fluorovaline presents a minor isosteric modification that allows continued recognition of the amino acid by the *E. coli* protein synthesis machinery. In the case of PpiB, the structure of the protein is minimally altered compared to the wild-type protein, even though valine constitutes almost 10% of the amino acid sequence. Owing to the exceptionally large chemical shift dispersion delivered by the γ-gauche effect, fluorovaline constitutes a highly attractive probe for site-specific interrogation of a target protein by 1D ^{19}F NMR spectra. *In vivo* systems have successfully been developed for genetic encoding of aromatic fluorinated amino acids.[37–45] Genetic encoding of fluorovaline could deliver a probe for inconspicuous installation at many more different sites of a protein.

UniProt ID

E. coli protein peptidyl–prolyl *cis–trans* isomerase B (PpiB): UniProt ID P23869

Supporting Information

Intact protein mass spectrum of PpiB-FVal; comparison of the thermal stability of PpiB-FVal and wild-type PpiB; electron densities in chains A–D indicating a conformation change in the side chain of Phe4; spatial distribution of the fluorinated valine residues throughout PpiB-FVal; table of the crystallographic statistics of PpiB-FVal; table of rotamer conformations of FVal residues in the crystal structure; and table of computed rotamer energies, ^{19}F chemical shifts, and $^3J_{HF}$ coupling constants

Acknowledgments

We thank Prof. Martin Scanlon for access to the 500 MHz NMR spectrometer. This 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, including the Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and a Discovery Project (DP230100079) is gratefully acknowledged.

References

(1) Gerig, J. T. Fluorine NMR of proteins. *Prog. NMR Spectrosc*. **1994**, *26*, 293–370.

(2) 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.

(3) Murray-Rust, P.; Stallings, W. C.; Monti, C. T.; Preston, R. K.; Glusker, J. P. Intermolecular interactions of the C−F bond: the crystallographic environment of fluorinated carboxylic acids and related structures. *J. Am. Chem. Soc.* **1983**, *105*, 3206−3214.

(4) Howard, J. A. K.; Hoy, V. J.; O'Hagan, D.; Smith, G. T. How good is fluorine as a hydrogen bond acceptor? *Tetrahedron* **1996**, *52*, 12613–12622.

(5) Dunitz, J. D. Organic fluorine: odd man out. *ChemBioChem* **2004**, *5*, 614–621.

(6) Dalvit, C.; Vulpetti, A. Weak intermolecular hydrogen bonds with fluorine: detection and implications for enzymatic/chemical reactions, chemical properties, and ligand/protein fluorine NMR screening. *Chem. Eur. J*. **2016**, *22*, 7592–7601.

(7) Tan, Y. J.; Abdelkader, E. H.; Tarcoveanu, E.; Maleckis, A.; Nitsche, C. and Otting, G. (2024) (2*S*,4*S*)-5-Fluoroleucine, (2*S*,4*R*)-5-fluoroleucine, and 5,5'-difluoroleucine in *E. coli* PpiB: protein production, ^{19}F NMR, and ligand sensing enhanced by the γ-gauche effect. *Biochemistry* **2024**, *63*, 1376–1387.

(8) Frkic, R. L.; Yan, Y. J.; Abdelkader, E. H.; Maleckis, A.; Tarcoveanu, E.; Nitsche, C.; Otting, G.; Jackson, C. J. Conformational preferences of the non-canonical amino acids(2*S*,4*S*)-5-fluoroleucine, (2*S*,4*R*)-5-fluoroleucine, and 5,5'-difluoroleucine in a protein. *Biochemistry* **2024**, *63*, 1388–1394.

(9) Maleckis, A.; Abdelkader, E. H.; Herath, I. D.; Otting, G. Synthesis of fluorinated leucines, valines and alanines for use in protein NMR. *Org. Biomol. Chem*. **2022**, *20*, 2424–2432.

(10) Feeney, J.; McCormick, J. E.; Bauer, C. J.; Birdsall, B.; Moody, C. M.; Starkmann, B. A.; Young, D. W.; Francis, P.; Havlin, R. H.; Arnold, W. D.; Oldfield, E. ^{19}F nuclear magnetic resonance chemical shifts of fluorine containing aliphatic amino acids in proteins: studies on *Lactobacillus casei* dihydrofolate reductase containing (2*S*,4*S*)-5-fluoroleucine. *J. Am. Chem. Soc*. **1996**, *118*, 8700–8706.

(11) Aragão, D.; Aishima, J.; Cherukuvada, H.; Clarken, R.; Clift, M.; Cowieson, N. P.; Ericsson, D. J.; Gee, C. L.; Macedo, S.; Mudie, N. MX2: a high-flux undulator

microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian synchrotron. *J. Synchr. Rad.* **2018,** *25*, 885–891.

(12) Kabsch, W. Xds. *Acta Cryst. D* **2010,** *66*, 125–132.

(13) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Cryst. D* **2013,** *69*, 1204–1214.

(14) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40*, 658–674.

(15) 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.

(16) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

(17) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Cryst. D* **2009**, *65*, 1074–1080.

(18) Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D. Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **2008,** *426*, 419–435.

(19) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst. D* **2004,** *60*, 2126–2132.

(20) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B. 3rd; Snoeyink, J.; Adams, P.D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*. **2018**, *27*, 293–315.

(21) Najibi, A.; Goerigk, L. The nonlocal kernel in van der Waals density functionals as an additive correction: an extensive analysis with special emphasis on the B97M-V and ωB97M-V approaches. *J. Chem. Theory Comput*. **2018**, *14*, 5725–5738.

(22) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys*. **2010**, *132*, 154104.

(23) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem*. **2011**, *32*, 1456–1465.

(24) Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron-affinities of the 1st-row atoms revisited – systematic basis-sets and wave-functions. *J. Chem. Phys*. **1992**, *96*, 6796–6806.

(25) Neese, F. An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix. *J. Comp. Chem*. **2003**, *24*, 1740–1747.

(26) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys*. **2009**, *356*, 98–109.

(27) Helmich-Paris, B.; de Souza, B.; Neese, F.; Izsák, R. An improved chain of spheres for exchange algorithm. *J. Chem. Phys*. **2021**, *155*, 104109.

(28) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic generation of auxiliary basis sets. *J. Theo. Comp. Chem*. **2017**, *13*, 554–562.

(29) Neese, F. Software update: the ORCA program system, version 5.0. *WIRES Comput. Molec. Sci*. **2022**, *12*, e1606.

(30) Neese, F. The SHARK integral generation and digestion system. *J. Comp. Chem*. **2022**, 1–16.

(30) Garcia-Rates, M.; Neese, F. Effect of the solute cavity on the solvation energy and its derivatives within the framework of the Gaussian charge scheme. *J. Comput. Chem*. **2020**, *41*, 922–939.

(31) Stoychev, G. L.; Auer, A. A.; Izsak, R.; Neese, F. Self-consistent field calculation of nuclear magnetic resonance chemical shielding constants using gauge-including atomic orbitals and approximate two-electron integrals. *J. Chem. Theory Comput*. **2018**, *14*, 619–637.

(32) Apponyi, M.; Ozawa, K.; Dixon, N. E.; Otting, G. *Methods in Molecular Biology* **2008**, *426*, Structural proteomics: high-throughput methods; B. Kobe, M. Guss, T. Huber Eds., Humana Press, Totowa, USA, pp. 257–268.

(33) Williamson, K. L.; Hus, Y.-F. L.; Hall, F. H.; Swager, S.; Coulter, M. S. Dihedral angle and bond angle dependence of vicinal proton–fluorine spin–spin coupling. *J. Am. Chem. Soc.* **1968**, *90*, 6717–6722.

(34) Gopinathan, M. S.; Narasimhan, P. T. Finite perturbation molecular orbital approach to the dihedral angle dependence of vicinal H-H and H-F couplings. *Mol. Phys*. **1971**, *21*, 1141–1144.

(35) Roise, D.; Soda, K.; Yagi, T.; Walsh, C. T. Inactivation of the *Pseudomonas striata* broad specificity amino acid racemase by D and L isomers of β-substituted alanines: kinetics, stoichiometry, active site peptide, and mechanistic studies. *Biochemistry* **1984**, *23*, 5195–5201.

(36) Kitevski-LeBlanc, J.-L.; Evanics, F.; Prosser, R. S. Optimizing ^{19}F NMR protein spectroscopy by fractional biosynthetic labeling. *J. Biomol. NMR* **2010**, *48*, 113–121.

(37) Furter, R. Expansion of the genetic code: site-directed *p*-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.* **1998**, *7*, 419–426.

(38) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **2011**, *50*, 1894−1900.

(39) Ko, J.; Wang, Y.-S.; Nakamura, A.; Guo, L.-T.; Söll, D.; Umehara, T. Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett.* **2013**, *587*, 3243–3248.

(40) Wang, Y.-S.; Fang, X.; Chen, H.-Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R. Genetic incorporation of twelve *meta*-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol*. **2013**, *8*, 405–415.

(41) Lee, Y.-J.; Schmidt, M. J.; Tharp, J. M.; Weber, A.; Koenig, A. L.; Zheng, H.; Gao, J.; Waters, M. L.; Summerer, D.; Liu, W. R. Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chem. Commun*. **2016**, *52*, 12606–12609.

(42) Qianzhu, H.; Abdelkader, E. H.; Herath, I. D.; Otting, G.; Huber, T. Site-specific incorporation of 7-fluoro-L-tryptophan into proteins by genetic encoding to monitor ligand binding by ^{19}F NMR spectroscopy. *ACS Sens.* **2022**, *7*, 44–48.

(43) Galles, G. D.; Infield, D. T.; Clark, C. J.; Hemshorn, M. L.; Manikandan, S.; Fazan, F.; Rasouli, A.; Tajkhorshid, E.; Galpin J. D.; Cooley, R. B.; Mehl, R. A.; Ahern, C. A. Tuning phenylalanine fluorination to assess aromatic contributions to protein function and stability in cells. *Nat. Commun*. **2023**, *14*, 59.

(44) Zhang, Q.; Zheng, W.; Song, Z.; Zhang, Q.; Yang, L.; Wu, J.; Lin, J.; Xu, G.; Yu, H. Machine learning enables prediction of pyrrolysyl-tRNA synthetase substrate specificity. *ACS Synth. Biol*. **2023**, *12*, 2403–2417.

(45) Qianzhu, H.; Abdelkader, E. H.; Otting, G.; Huber, T. Genetic encoding of fluoro-L-tryptophans for site-specific detection of conformational heterogeneity in proteins by NMR spectroscopy. *J. Am. Chem. Soc.* **2024**, *146*, 13641–13650.

For Table of Contents use only