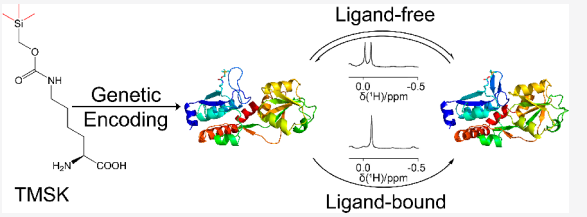
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Genetic Encoding of N^6-(((Trimethylsilyl)methoxy)carbonyl)-L-lysine for NMR Studies of Protein–Protein and Protein–Ligand Interactions

Elwy H. Abdelkader, Haocheng Qianzhu, Yi Jiun Tan, Luke A. Adams, Thomas Huber,* and Gottfried Otting*

Cite This: https://dx.doi.org/10.1021/jacs.0c11971
Read Online

ABSTRACT: Trimethylsilyl (TMS) groups present outstanding NMR probes of biological macromolecules as they produce intense singlets in ^{1}H NMR spectra near 0 ppm, where few other proton resonances occur. We report a system for genetic encoding of N^6-(((trimethylsilyl)-methoxy)carbonyl)-L-lysine (TMSK) for site-specific incorporation into proteins. The system is based on pyrrolysyl-tRNA synthetase mutants, which deliver proteins with high yield and purity *in vivo* and in cell-free protein synthesis. As the TMS signal can readily be identified in 1D ^{1}H NMR spectra of high-molecular weight systems without the need of isotopic labeling, TMSK delivers an excellent site-specific NMR probe for the study of protein structure and function, which is both inexpensive and convenient. We demonstrate the utility of TMSK to detect ligand binding, measure the rate of conformational change, and assess protein dimerization by paramagnetic relaxation enhancement. In addition, we present a system for dual incorporation of two different unnatural amino acids (TMSK and *O*-*tert*-butyl-tyrosine) in the same protein in quantities sufficient for NMR spectroscopy. Close proximity of the TMS and *tert*-butyl groups was readily detected by nuclear Overhauser effects.



INTRODUCTION

Engineering of pyrrolysyl-tRNA synthetase/pyrrolysyl-tRNA (PylRS/tRNA^{Pyl}) pairs has delivered systems for the genetic encoding of a wide variety of unnatural amino acids through stop codon suppression.[1] Despite the potential of the PylRS system for studying proteins by site-specific incorporation of unnatural amino acids, the protein yields obtained with most of the PylRS variants are generally modest, which has limited their application in biomolecular NMR.[2,3]

Here we report the site-specific incorporation of N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine (TMSK) *in vivo* and *in vitro* in high yield and demonstrate multiple uses in NMR spectroscopy of proteins. TMSK is a lysine derivative with a trimethylsilyl (TMS) group attached to N^6 (Figure 1). Using our recently reported library screening approach based on fluorescence-activated cell sorting (FACS),[4] a specific pyrrolysyl-tRNA synthetase was identified that recognizes TMSK. In conjunction with an optimized two-plasmid system, TMSK is incorporated into proteins in response to an amber stop codon with both high yield and fidelity.

The ^{1}H NMR signal of the TMS group is intense owing to nine equivalent protons that produce a singlet near 0 ppm, where interference from background protein signals is minimal. *p*-TMS-phenylalanine (TMSf) has similarly been incorporated into proteins previously, either using a polyspecific[5] or optimized tRNA-synthetase (RS) enzymes,[6] but aromatic ring currents imply that the intrinsic chemical shift of the

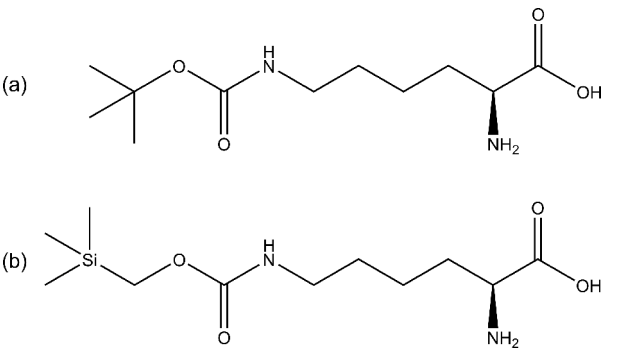


Figure 1. Chemical structures of (a) N^6-(*tert*-butoxycarbony)-L-lysine (BocK) and (b) N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine (TMSK).

TMS group in TMSf is at 0.13 ppm, i.e. slightly above 0 ppm. Moreover, the side chain of TMSK is longer and more flexible than that of TMSf, suggesting that the TMS group in TMSK

Received: November 15, 2020

can be governed by a shorter rotational correlation time than in TMSf. Flexible cysteine-reactive TMS tags have indeed been demonstrated to yield readily observable signals in large proteins,[7,8] but relying on ligation with cysteine residues limits the scope of this method, if naturally occurring cysteines are necessary for the structure or function of the protein.

Similar to TMS groups, *tert*-butyl groups have been used earlier for studying proteins by ^{1}H NMR.[5,8,9] A particularly straightforward way of introducing a *tert*-butyl group into proteins is by site-specific incorporation of *O*-*tert*-butyltyrosine (Tby).[9] Although the nine methyl protons of Tby produce a narrow and intense ^{1}H NMR signal, this signal appears at about 1.3 ppm and is thus more difficult to distinguish from the NMR signals of the protein than the signal of the TMS group.

In the following, we introduce new tRNA synthetases which genetically encode TMSK efficiently and with high specificity. We show how genetically encoded TMSK residues can be used to measure ligand binding affinities, detect dimerization with the help of paramagnetic labeling, and measure conformational exchange kinetics. In addition, we demonstrate a cell-free expression system for high-yield incorporation of TMSK. Finally, we demonstrate the simultaneous site-specific incorporation of Tby and TMSK into a two-domain protein to document an interdomain contact by a nuclear Overhauser effect (NOE) between the *tert*-butyl and TMS groups.

RESULTS

Site-Specific Incorporation of TMSK into Proteins. In view of the structural similarity between TMSK and BocK (Figure 1), which is known to be readily recognized by the wild-type PylRS/tRNA^{Pyl} pair from *Methanosarcina mazei* (*Mm*), we initially attempted incorporation of TMSK by the same *Mm* PylRS/tRNA^{Pyl} pair. This was unsuccessful. Similarly, TMSK was recognized neither by the PylRS Y384F mutant[10] nor the Y306G/Y384F/I405R triple mutant, which can recognize lysine derivatives with large terminal groups.[11]

We therefore produced a library of PylRS mutants to select a TMSK-recognizing mutant, using our previously published library screening approach based on fluorescent activated cell sorting (FACS)[4] enabled by the expression of mCherry red fluorescent protein (RFP) preceded by a His_6 tag and an amber stop codon (His_6-TAG-RFP) from the selection plasmid pBAD-H6RFP. The library was based on a chimeric PylRS variant, consisting of residues 1−149 of the *Methanosarcina barkeri* (*Mb*) PylRS (*Mb*PylRS) and residues 185−454 of *Mm*PylRS. This chimeric RS enzyme was reported to be more soluble than the parent wild-type synthetases. In addition, the C-terminal *Mm*PylRS domain contained the mutation Y384F (in the sequence numbering of the wild-type enzyme) reported to increase the aminoacylation activity of PylRS.[10] The final library plasmid pBK-ChPylRS encoded the ChPylRS mutants, where the residues A302, L305, Y306, N346, C348, and V401 of the substrate binding pocket were fully randomized, and an orthogonal $^{Mb}tRNA_{CUA}^{Pyl}$ in a pBK vector.[4] Another selection plasmid, pBAD-H6RFP, contained a mCherry red fluorescent protein (RFP) gene following a His_6-tag and amber stop codon. The library was cotransformed with the plasmid pBAD-H6RFP into *E. coli* DH10B for selection via FACS. After six rounds of selection (three positive rounds with TMSK and three negative rounds without), separate samples of the remaining library displayed a dramatic difference in RFP fluorescence in the presence and absence of TMSK (Figures S4 and S5), indicating successful enrichment of the desired ChPylRS variants in the pool of the remaining library. Eight clones producing the highest RFP fluorescence were sequenced and their sequences converged to two different mutants: C348G/V401C/Y384F (ChPylTMSK) and V401K/Y384F. The specific incorporation of TMSK by these two ChPylRS mutants was confirmed by mass spectrometry analysis.

Optimization of the Two-Plasmid System for *In Vivo* TMSK Incorporation. Following the cloning of ChPylTMSK into a pEVOL plasmid[12] modified to contain the orthogonal $^{Mb}tRNA_{CUA}^{Pyl}$, the resulting plasmid pEVOL-ChPylTMSK was used to test the efficiency of TMSK incorporation into proteins in the *E. coli* BL21(DE3) overexpression strain. A pET3a plasmid coding for the expression of an amber mutant of *E. coli* peptidyl-prolyl *cis-trans* isomerase B (PpiB-H147TAG) was cotransformed and cells were grown in LB media containing 1 mM TMSK, 0.2% L-arabinose, and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Disappointingly, only about 1 mg of purified PpiB mutant per liter cell culture was obtained and mass spectrometry revealed an amber suppression efficiency of less than 10% (Figure S6a). We therefore tested alternative plasmid combinations to enhance the protein yield and fidelity of TMSK incorporation (Table S1).

The best system carried the PylRS/tRNA^{Pyl} pair on a pRSF plasmid[13] under control of T7 and *proK* promoters, respectively, and the gene for the PpiB H147TAG mutant on a pCDF plasmid[14] under control of a T7 promoter (Table S2). The new pRSF/pCDF system yielded about 8 mg purified PpiB H147TMSK per liter of cell culture and intact protein mass spectrometry indicated >95% TMSK incorporation (Figure S6b). A further 3-fold increase in yield without any reduction in fidelity was obtained by using the pRSF/pCDF system in the release factor 1 (RF-1) knockout *E. coli* BL21(DE3) strain B-95.ΔA[15] (Figure S6c).

Figure 2 shows the 1D ^{1}H NMR spectra of PpiB H147TMSK and wild-type PpiB. The nine protons of the TMS group appear as a single intense peak at about 0.06 ppm.

Incorporation of TMSK into the *Bacillus stearothermophilus* (*Bst*) DnaB Hexamer. Using the same optimized system, TMSK was successfully incorporated into *Bst* DnaB at position Y104 (yielding 10 mg purified protein per liter of cell culture; the construct was an N-terminal fusion with GB1 to enhance the expression yields).[9] The hexameric state (320 kDa) was confirmed by gel filtration. Despite the high molecular weight and loss of protein during gel filtration, the signal of the TMS group of DnaB Y104TMSK was clearly visible in the 1D ^{1}H NMR spectrum (Figure 3). The unique chemical shift of the TMS resonance near 0 ppm renders its assignment straightforward. Previous experiments with *tert*-butyl-tyrosine in *Bst* DnaB required protein perdeuteration to unambiguously confirm the assignment of the *tert*-butyl signal, which appeared at about 1.2 ppm, to avoid potential confusion with minor impurities.[9]

TMSK for the Measurement of Dissociation Constants. The ease with which the TMS group of TMSK can be detected in 1D ^{1}H NMR spectra of high-molecular weight systems invites its use as a probe to detect site-specific changes, such as effected by ligand binding. As an example, we incorporated TMSK at sites E28, N117, or H160 of the widely used polysubstrate-specific *p*-cyano-L-phenylalanyl-tRNA synthetase (CNRS) derived from *Methanocaldococcus jannaschii* (*Mj*) tyrosyl−tRNA synthetase, which is a homodimeric

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

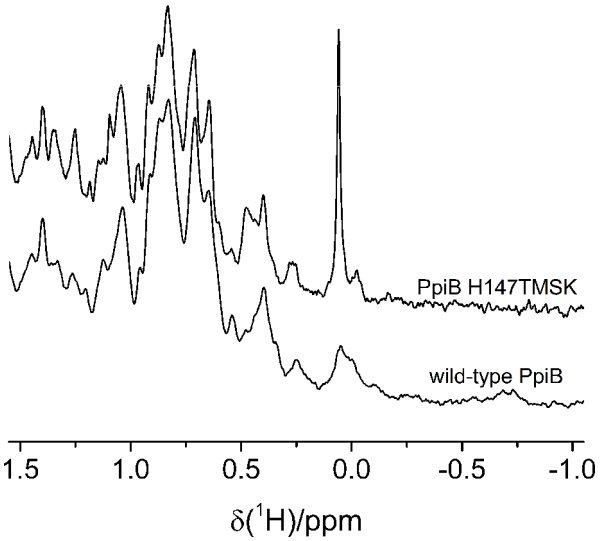


Figure 2. 1D ^{1}H NMR spectra of wild-type PpiB and its mutant H147TMSK. The spectra were recorded in a buffer of 50 mM HEPES-KOH, pH 7.5, at 25 °C using a Bruker 600 MHz NMR spectrometer equipped with a TCI cryoprobe. Protein concentrations were about 60 μM, and the total recording time per spectrum was 12 s.

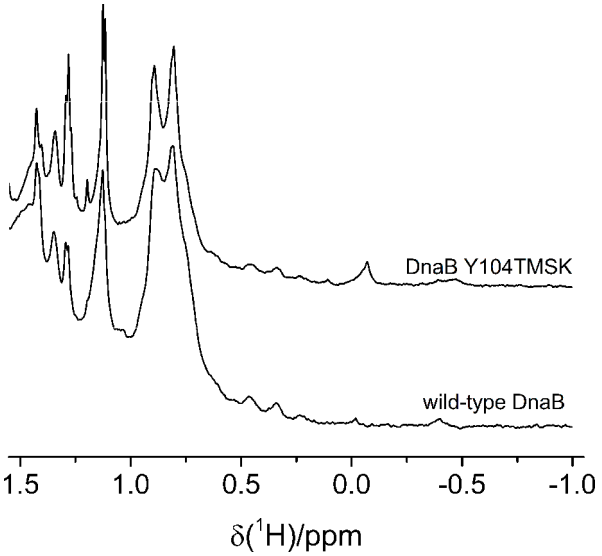


Figure 3. 1D ^{1}H NMR spectra of a 100 μM solution of *Bst* DnaB and 25 μM solution of its mutant Y104TMSK. The spectra were recorded in 25 mM Tris-HCl, pH 7.5, 150 mM NaCl, at 25 °C using a Bruker 600 MHz NMR spectrometer. The total recording time per spectrum was 4 and 30 min, respectively.

enzyme of 71 kDa molecular weight.[16] Based on the crystal structure (PDB ID: 3QE4), the backbone atoms of the residues in positions 28, 117, and 160 are within 7−18 Å of the *p*-cyano-L-phenylalanine (CNF) residue in the active site (Figure 4).

The ^{1}H NMR signal of the TMS group was readily observable between −0.04 and −0.01 ppm for all three mutants (Figure 5). The line widths of the signals decreased with increasing distance of the mutation site from the dimer interface. For TMSK in position 117, the TMS resonance observed at 800 MHz was heterogeneous and could be deconvoluted into two signals of equal area but different width,

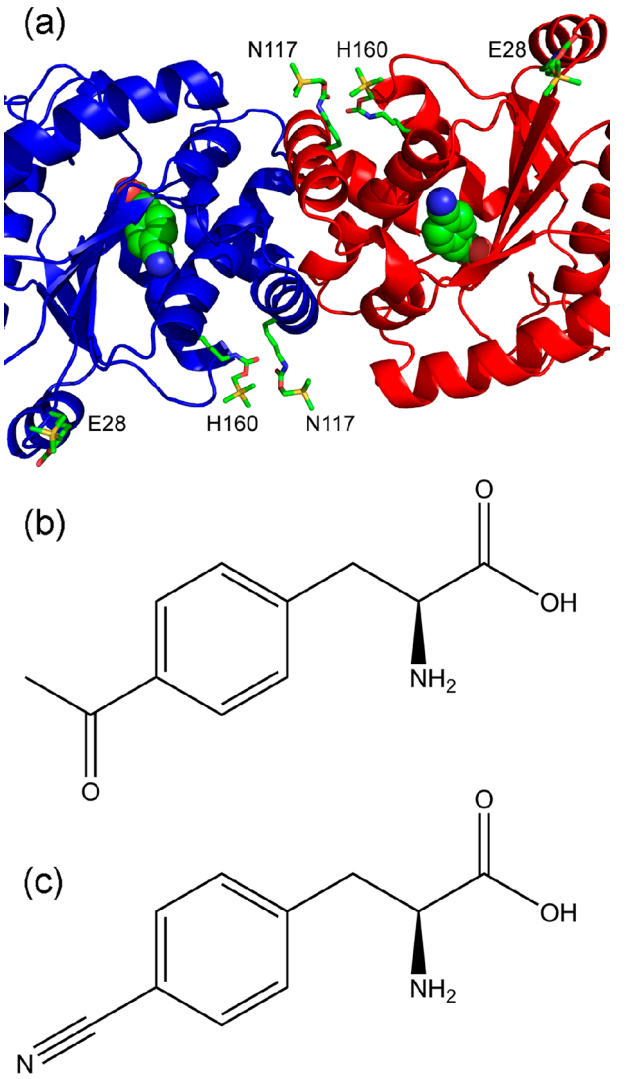


Figure 4. Homodimer interface of CNRS with locations of mutation sites and chemical structures of amino acid ligands. (a) Ribbon representation of the homodimer interface of CNRS (PDB: 3QE4)[16] showing the monomeric subunits in blue and red, respectively. TMSK residues modeled at the positions indicated are shown in stick representations. CNF in the active sites is highlighted by spheres. (b) Chemical structures of *p*-acetyl-L-phenylalanine (AcF) and (c) *p*-cyano-L-phenylalanine (CNF).

whereas the TMS resonances in the mutants E28TMSK and H160TMSK appeared as singlets. The difference in chemical environments experienced by the TMS groups in the CNRS N117TMSK dimer appeared to vanish upon titration with the amino acids CNF and *p*-acetyl-L-phenylalanine (AcF) (Figures 4b, c and 6), which are known to bind to the active site of the enzyme. These observations could be explained by some asymmetry in the dimer interface in the absence, but not in the presence, of amino acid. Notably, position 117 is closer to the dimer interface than the positions 28 and 160 (Figure 4a).

Besides rendering the line shape of the TMS resonance more homogeneous, titration of CNRS N117TMSK with AcF or CNF also led to small chemical shift changes of the TMS resonance (Figure 6), which could be used to assess the relative binding affinities. Although CNRS is polyspecific,[16] it had originally been selected for activity with CNF.[17] Consistent with this, CNRS N117TMSK was found to bind

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

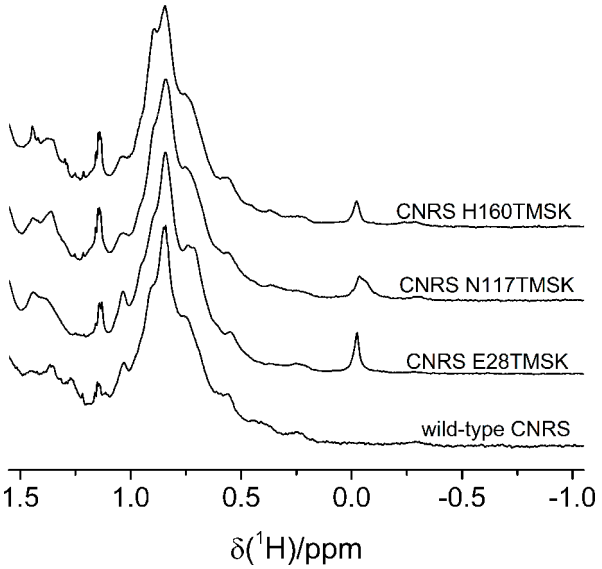


Figure 5. 1D ^{1}H NMR spectra of CNRS E28TMSK, N117TMSK, H160TMSK, and wild-type CNRS. The spectra were recorded 100 μM solutions in buffer A (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM adenylyl-imidodiphosphate (AMP-PNP), 1 mM dithiothreitol (DTT)) at 25 °C on a 800 MHz NMR spectrometer. The total recording time per spectrum was about 4 min.

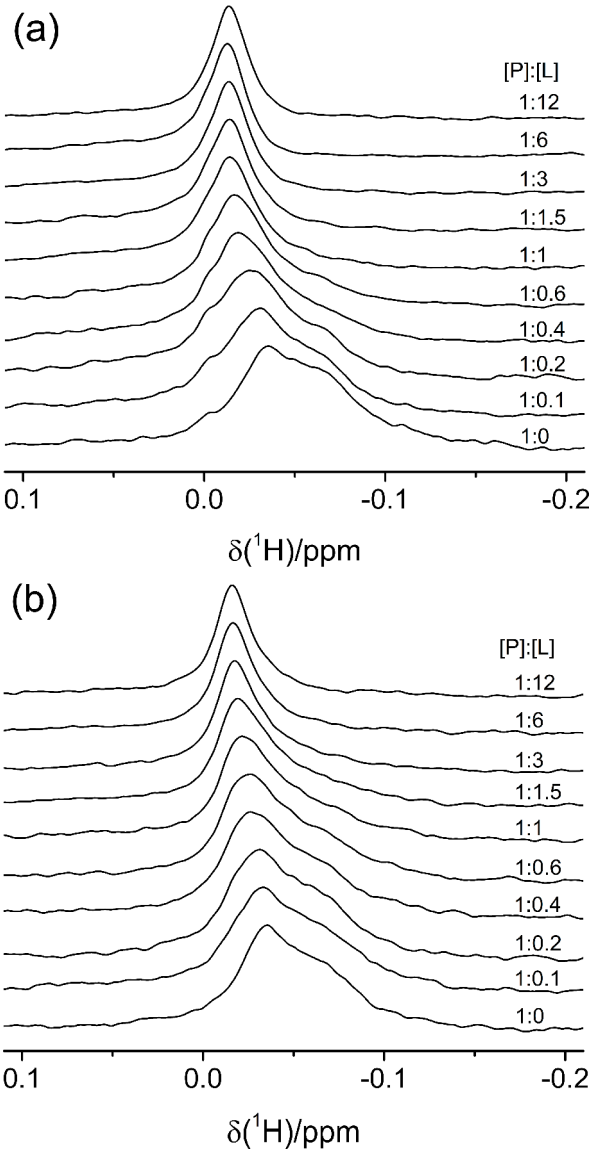


Figure 6. Titration of CNRS N117TMSK in NMR buffer with (a) CNF and (b) AcF. The spectra were recorded at 25 °C on a 800 MHz NMR spectrometer in about 4 min each. The protein concentration [P] was 100 μM and the ligand concentration [L] was as indicated.

CNF more tightly than AcF. Following the peak maximum of the TMS signal, a 1:1 binding model suggested dissociation constants K_d in the low micromolar range, but the fit was dissatisfactory especially with CNF (Figure 7a and b). Exploring the cooperativity of binding to the two substrate binding sites in the dimer by fitting the Hill equation (eq S2), the titration data with CNF and AcF yielded apparent K_d values of about 25 and 50 μM, respectively (Figure 7c), with positive Hill coefficients. This suggests cooperative binding to the two monomers, reflecting the dimeric state of the enzyme.

Tighter binding of CNF compared with AcF was also observed with CNRS H160TMSK. In this case, titration with either amino acid led to the disappearance of the original TMS signal and the appearance of a new peak at 0.013 ppm for ligand-saturated CNRS, indicating slow exchange between CNRS with and without bound amino acid (Figure S8). Deconvoluting the signal into its two components and using eq S1 to fit the peak integrals indicated dissociation constants of 9 and 14 μM for CNF and AcF, respectively. Also in this case, the Hill equation delivered a better fit (Figures S9 and S10).

Titration of CNRS E28TMSK with CNF indicated a weaker binding affinity than the other two TMSK mutants (Figure S11). The chemical shift of the TMS resonance changed gradually during the titration, indicating that the exchange was fast on the NMR time scale. The apparent dissociation constant determined with the Hill equation was 114 μM, with a Hill coefficient of 2.2. The weaker binding affinity observed for CNRS E28TMSK is an example of an increasing body of evidence that enzymatic function can be modulated by solvent-exposed residues distant from the substrate binding site.[18] In contrast to the other two mutations, the E28TMSK mutation replaced a negatively charged residue by an uncharged residue.

Probing Protein–Protein Interfaces with TMSK and Paramagnetic Relaxation Enhancements (PREs). To probe the dimerization interface in the CNRS homodimer, CNRS N117TMSK was mixed with a CNRS mutant tagged with a paramagnetic tag in position 81. To allow usage of a cysteine-reactive tag, the naturally occurring cysteine residues 190 and 231 were mutated to serine and Glu81 was mutated to cysteine. This protein was labeled with the paramagnetic C1-Gd^{3+} or diamagnetic C1-Y^{3+} tag[19] (in the following referred to as CNRS-Gd and CNRS-Y, respectively). The 1D NMR spectrum of an equimolar mixture of CNRS N117TMSK and CNRS-Gd showed a reduced intensity of the TMS signal compared to the diamagnetic reference constituted of CNRS N117TMSK and CNRS-Y (Figure 8). The observed halving of the TMS signal in the paramagnetic sample is expected for random association between the TMSK and C1-Gd^{3+} labeled proteins, when the TMS signal is broadened beyond detection by the Gd^{3+} ion due to paramagnetic relaxation enhancement (PRE). Repeating the same experiment with the CNRS C190S/C231S/E81C mutant tagged with the nitroxide tag MTSL produced no change in the TMS peak height, indicating that the paramagnetism of the nitroxide group was insufficient

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

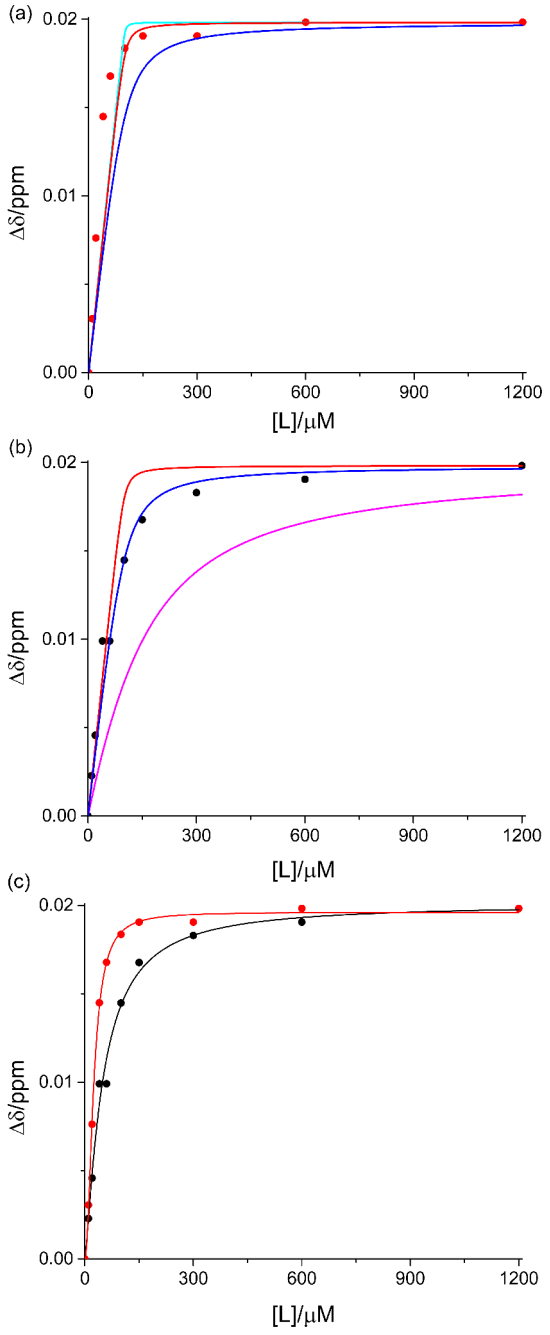


Figure 7. Binding affinities of CNRS N117TMSK for CNF and AcF. (a) Chemical shift change of the maximum of the TMS signal in response to titration with CNF, displayed together with binding curves expected for 1:1 binding (eq S1 in the Supporting Information) with dissociation constants of 0.1 (cyan), 1 (red), and 10 μM (blue), respectively. (b) Same as part a, but for titration with AcF. The magenta curve corresponds to K_d = 100 μM. (c) Titration data of parts a and b fitted with the Hill equation (eq S2 in the Supporting Information). The fits yielded apparent dissociation constants K_d of, respectively, 49 and 24 μM with Hill coefficients of 1.3 and 2 for the titrations with CNF (red data) and AcF (black data), respectively.

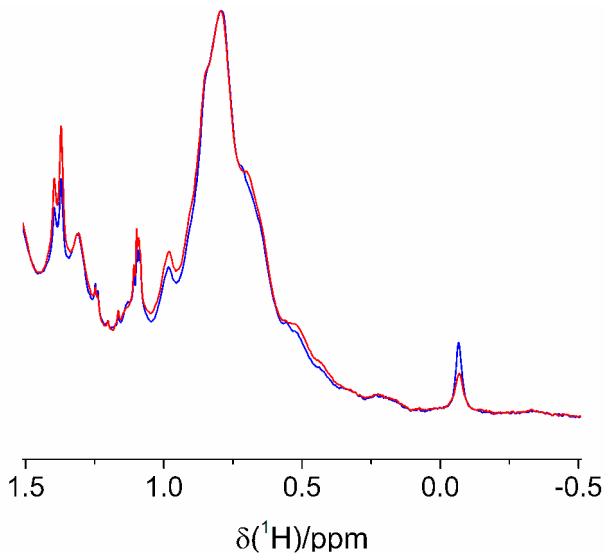


Figure 8. 1D ^{1}H NMR spectra of CNRS N117TMSK mixed in equimolar ratio with paramagnetic CNRS-Gd (red spectrum) or diamagnetic CNRS-Y (blue spectrum). The spectra were recorded in a buffer of 25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM AMP-PNP, and 1.2 mM CNF at 25 °C using a 800 MHz NMR spectrometer. Protein concentrations were 100 μM each, and each spectrum was recorded in 4 min.

to elicit a significant PRE. This observation can be interpreted in terms of the distance between the positions 81 and 117.

For a fixed distance r between a nitroxide and a ^{1}H spin in the slow tumbling limit, the paramagnetic enhancement of transverse relaxation, Γ_2, can be translated into a distance using the equation[20]

$$\Gamma_2 = C \times S(S+1)\tau_c/r^6 \quad (1)$$

with $C = 6.5 \times 10^{16}$ Å^6 s^{-2}. Assuming an electronic spin S = 1/2, an effective correlation time τ_c = 3 ns (i.e., shorter than the rotational correlation time of the protein dimer to account for the flexibility of the TMSK side chain as well as the cysteine side chain with MTSL tag) and that Γ_2 = 300 s^{-1} would broaden the ^{1}H NMR signal beyond detection, disappearance of the TMS resonance is expected for r < 9 Å. For the PRE from a Gd^{3+} ion, for which S = 7/2, the corresponding cutoff distance would be r < 15 Å. Using both the MTSL and gadolinium tags thus establishes an approximate distance restraint in this distance range. Modeling a distance distribution by rotamer libraries of TMSK and the C1-Gd^{3+} tag crafted onto the crystal structure of CNRS suggested 21 Å as the most frequent distance between the methyl protons of the TMS group and the Gd^{3+} ion, but due to the steep distance dependence of the PRE, the biggest contribution is expected to be from protons that are about 11 Å from the Gd^{3+} ion (Figure S12). This is less than 15 Å, for which the Gd^{3+} tag is expected to broaden the TMS resonance beyond detection.

Modeling the equivalent distance distribution for the sample labeled with MTSL indicates that the most frequent distance of the paramagnetic center to the TMS protons similarly is about 21 Å, but the biggest contribution to the PRE is expected from TMS protons that are about 14 Å from the nitroxide group (Figure S12). This is longer than the maximal distance of 9 Å, for which the PRE is expected to be substantial and explains the difficulty to observe the PRE with MTSL in our experiments. The different linker lengths and conformations of the C1-Gd^{3+} and MTSL tags thus contribute less to their difference in PRE effects than their different electronic spins.

NOEs with the TMS Resonance. Due to its position at the very end of the flexible side chain of the TMSK amino acid,

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

the effective rotational correlation time of the TMS group is shorter than for backbone atoms of the protein, but this does not prevent the observation of NOEs with the TMS resonance. For example, the CNRS mutants N117TMSK and H160TMSK both showed a set of similar NOEs (Figure 9).

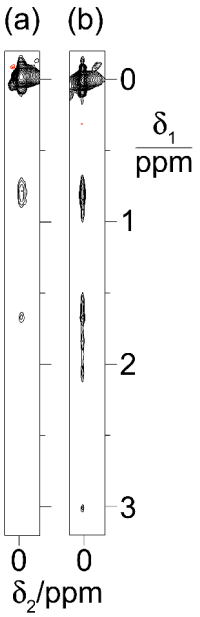


Figure 9. Selected spectral regions from NOESY spectra of TMSK mutants of CNRS, showing NOE cross-peaks with the TMS resonance. (a) Spectrum of a 1 mM solution of CNRS N117TMSK in 25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM AMP-PNP, 2 mM $MgCl_2$, and 2 mM DTT. The spectrum was recorded at 25 °C on a 600 MHz NMR spectrometer. Mixing time 200 ms, t_{1max} = 28 ms, t_{2max} = 142 ms, and total recording time 2.3 h. (b) Same as for part a, but for CNRS H160TMSK. The protein concentration was 0.6 mM, and the spectrum was recorded in 9 h.

We have not assigned these NOEs to specific protein resonances but note that chemical shifts below 1 ppm are characteristic of the methyl groups of valine, leucine and isoleucine. In the crystal structure (PDB ID 3QE4),[16] the side chains of these two residues are very close to each other and in contact with the isopropyl groups of Leu121 and Leu184. Despite the length of the TMSK side chains, it is conceivable that the TMS groups transiently form hydrophobic contacts with these solvent-exposed isopropyl groups. Indeed, if the TMS groups were only experiencing a fully hydrated state far from the protein surface, it would be difficult to explain the sensitivity of the TMS chemical shift to ligand binding elsewhere in the protein.

Cell-Free Protein Synthesis (CFPS) Incorporation of TMSK. With the identification of a new class of orthogonal PylRS enzymes lacking the N-terminal domain,[21] it has become possible to purify PylRS variants with high enzymatic activity, which can be added to CFPS reactions for efficient incorporation of unnatural amino acids.[22] Among this new class, the PylRS enzymes from *Methanomethylophilus alvus* (*Ma*PylRS) and the methanogenic archaeon ISO4-G1 (G1PylRS) were chosen to test their ability to incorporate TMSK in CFPS. To transfer the C348G/V401C/Y384F mutations of ChPylTMSK, the sequence alignment with *Mm*PylRS suggested that the equivalent set of mutations would be V168G/A223C/Y206F in *Ma*PylRS and V167G/A221C/Y204F in G1PylRS (Figure S13). We refer to these mutants as *Ma*PylTMSK and G1PylTMSK and tested their ability to recognize and incorporate TMSK *in vitro*.

*Ma*PylTMSK and G1PylTMSK were expressed *in vivo* and the purified enzymes together with the respective orthogonal $tRNA_{CUA}^{Pyl}$ were added to a CFPS system based on an RF-1-free S30 extract made from *E. coli* cells where RF-1 was tagged with chitin binding domains for selective removal by filtration through chitin.[23] The reporter protein was an RFP construct with an amber stop codon inserted after the start codon to produce RFP TMSK. Both *Ma*PylTMSK and G1PylTMSK significantly increased the RFP fluorescence level, when 1 mM TMSK was present in the CFPS reaction. Compared to wild-type RFP, *Ma*PylTMSK and G1PylTMSK delivered 20% and 65% amber suppression efficiency, respectively (Figure S14), and mass spectrometry of RFP TMSK purified from the CFPS reaction containing G1PylTMSK showed high fidelity of TMSK incorporation (Figure S15). Therefore, G1PylTMSK was chosen for TMSK incorporation in CFPS. Indeed, using CFPS to incorporate TMSK into the RFP construct yielded protein in sufficient amount for NMR spectroscopy (0.5 mg purified protein per 1 mL of CFPS inner reaction mixture, Figure 10).

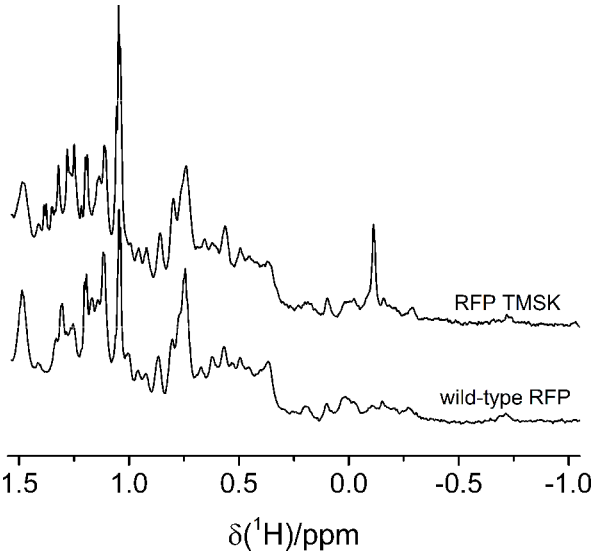


Figure 10. High-field spectral region of the 1D ^1H NMR spectra of wild-type RFP and RFP with a TMSK residue incorporated following the initiating methionine residue. The spectra were recorded in 25 mM Tris-HCl, pH 7.5, 150 mM NaCl at 25 °C using a 800 MHz NMR spectrometer. Protein concentrations were about 50 μM. The total recording time per spectrum was about 4 min.

Two Different Unnatural Amino Acids in the Same Protein. The mutual orthogonality of the PylRS/$tRNA^{Pyl}$ and *Mj* tyrosyl-tRNA synthetase/$tRNA_{UCA}^{Tyr}$ pairs opens the possibility to incorporate both TMSK and Tby in the same polypeptide chain in response to an amber and opal stop codon, respectively. In view of the hydrophobicity of the TMS and *tert*-butyl groups, we hypothesized that an NOE could be detected between both groups even when conformational freedom would suggest direct contacts to be rare. To test this hypothesis, we chose a two-domain amino acid binding protein known to undergo a conformational change between an open and closed conformation.

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

The ancestral protein of the cyclohexadienyl dehydratase (AncCDT-1)[24] is a monomeric protein that undergoes a conformational change upon binding to L-arginine to a closed conformation. In the absence of ligand, distance measurements by EPR spectroscopy have shown that the closed conformation coexists with an open conformation.[25] Based on crystal structures of both conformations, Pro57 in the N-terminal domain and Asn151 in the C-terminal domain were chosen for the incorporation of TMSK and Tby, respectively (Figure 11a and b).

The site-specific dual incorporation of TMSK and Tby was achieved by a three-plasmid system comprised of pRSF-ChPylTMSK (amber suppression), pEVOL-CNRS (opal suppression), and pCDF-AncCDT-1-P57TAG/N151TGA. AncCDT-1 P57TMSK/N151Tby was obtained in good yield (24 mg purified protein per liter of cell medium) and purity (Figure S16). Figure 11c compares the 1D ^{1}H NMR spectrum of AncCDT-1 P57TMSK/N151Tby with that of the wild-type protein. Two narrow singlets of similar intensity at −0.023 and 1.318 ppm correspond to the nine methyl protons of the TMS and *tert*-butyl groups, respectively.

Previous interdomain distance measurements had been conducted by double electron−electron resonance (DEER) measurements, using AncCDT-1 samples labeled with Gd^{3+}-tags at positions 68 and 138. Three different samples were investigated: (i) A single Gd^{3+}−Gd^{3+} distance distribution was observed for fresh samples. (ii) Two Gd^{3+}−Gd^{3+} distance distributions of similar abundance were detected following a cycle of denaturation and refolding to remove any ligand that was captured during protein expression and purification. The distances were characteristic of the open and closed conformations detected previously by X-ray crystallography. (iii) The addition of L-arginine to the refolded protein reproduced the distance distribution obtained for fresh samples, which corresponded to the closed conformation.[25] In agreement with these EPR results, the TMS group in AncCDT-1 P57TMSK/N151Tby displayed a single peak in the freshly purified protein, two peaks after denaturation and refolding, and a single peak again after addition of arginine (Figure 11c). Furthermore, a NOESY spectrum recorded of the refolded sample (Figure 12a) showed exchange cross-peaks between the two TMS signals corresponding to an exchange rate of 0.5 s^{-1} (Figure 12b). In addition, NOEs between the TMS and *tert*-butyl group were clearly observable for the high-field TMS signal, whereas the low-field TMS peak displayed an NOE to a different resonance, which was broader than the *tert*-butyl resonance. This NOE was observable only on one side of the diagonal, as the longer acquisition time in the F_2 dimension favored narrow over broad signals. The NOE of the high-field TMS peak gained in intensity after addition of arginine in proportion to the increased intensity of the merged TMS signals, as expected if the closed conformation is the same in the presence and absence of arginine (Figure 12c). The positive sign of the NOE cross-peaks indicates that the contact between the *tert*-butyl and TMS groups is in the slow motional regime, with a correlation time of the interaction of at least 0.3 ns.

The NOESY spectrum revealed a clear technical advantage of measuring cross-peaks of resonances near 0 ppm, where the protein displays few other resonances. This region in the spectrum is less affected by t_1 noise and much easier to baseline correct than the spectral region above 0.2 ppm, where

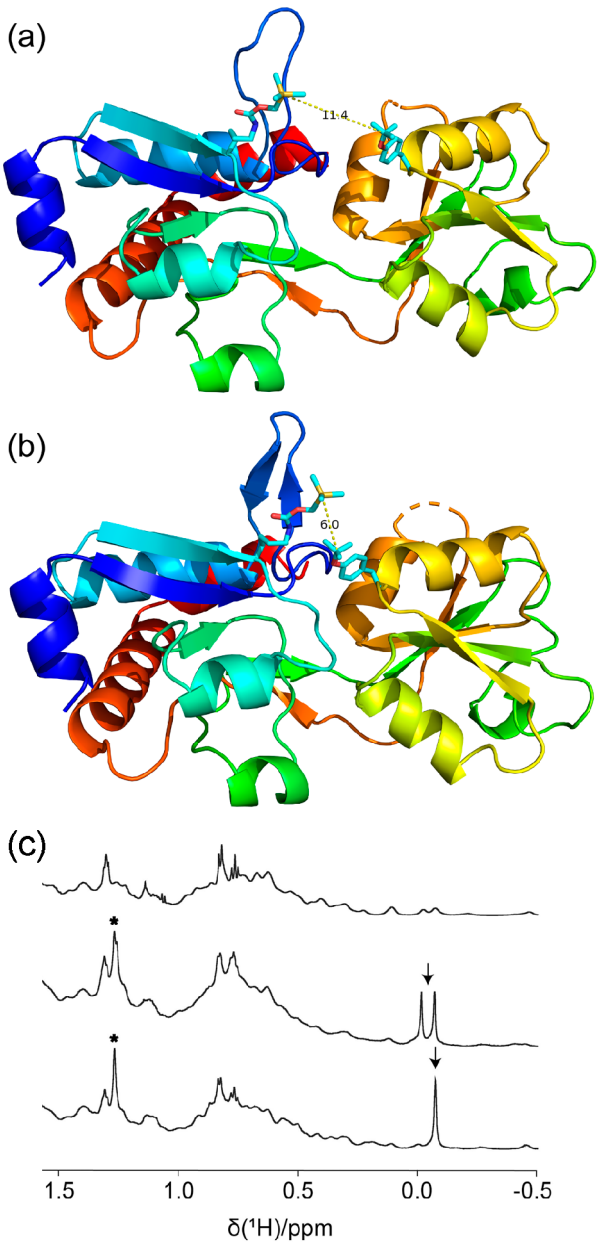


Figure 11. Dual incorporation of TMSK and *O*-*tert*-butyl-tyrosine (Tby) into a single protein. (a) Cartoon representation of AncCDT-1 modeled with TMSK and Tby residues in positions 57 and 151, respectively (side chains shown in a stick representation). The distances shown (in Å) are between the silicon atom of the TMS group and the quaternary carbon of the *tert*-butyl group. The figure shows the open conformation of AncCDT-1 (PDB: 5TUJ).[24] (b) Same as part a, but displaying the closed conformation (PDB: 5T0W). (c) 1D ^{1}H NMR spectra of wild-type AncCDT-1(top panel), AncCDT-1 P57TMSK/N151Tby (middle panel), and AncCDT-1 P57TMSK/N151Tby following addition of an equimolar amount of L-arginine (bottom panel). The spectra were recorded in 50 mM HEPES-KOH, pH 7.5, 150 mM NaCl at 25 °C using a 600 MHz NMR spectrometer. Protein concentrations were 70−140 μM, and the total recording time per spectrum was about 15 min. Note that, in the apoprotein, two signals were observed for the TMS group (labeled with an arrow) and two barely resolved singlets are also visible for the *tert*-butyl group of Tby (labeled with a star). Following addition of L-arginine, the peak heights of the remaining TMS and *tert*-butyl resonances increased relative to the protein signals, as expected for the transition from two conformations to a single conformation.

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

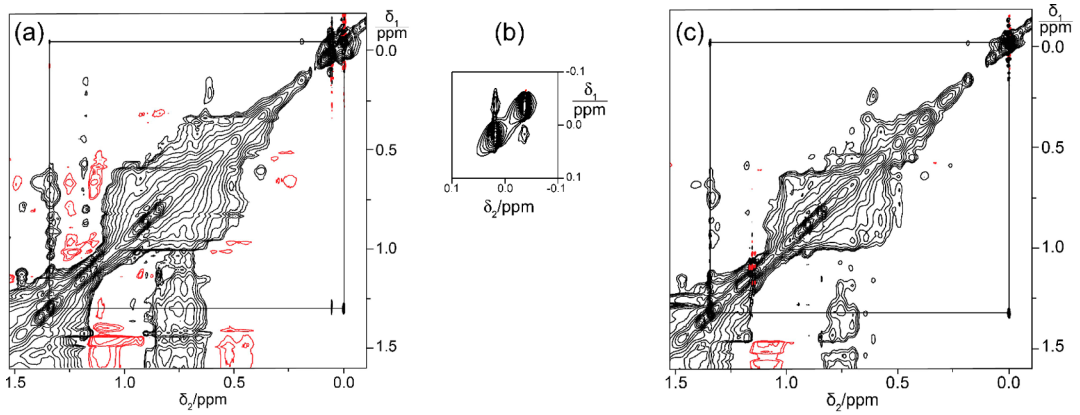


Figure 12. NOEs and exchange cross-peaks of the TMS resonance in a 0.6 mM solution of AncCDT-1 P57TMSK/N151Tby in 50 mM HEPES, pH 7.5, 150 mM NaCl at 25 °C, recorded with the pulse sequence of Figure S3. Jump-return delay 88 μs, 200 ms mixing time, t_{1max} = 25 ms, t_{2max} = 0.56 s, 600 MHz NMR spectrometer, total recording time 17 h. (a) 2D NOESY spectrum of the protein after denaturation and refolding. Vertical and horizontal lines drawn from the diagonal peaks of the *tert*-butyl and TMS groups meet at the cross-peaks between these two groups. (b) Spectral region of the TMS signals plotted at higher contour levels to highlight the exchange cross-peaks. (c) Same as part a but after addition of an equimolar amount of L-arginine.

the diagonal peaks are taller and there is little baseline to guide the correction algorithm.

Comparison of the NOESY spectra of the single-TMSK mutants of CNRS (Figure 9) with those of the P57TMSK/N151Tby mutant (Figure 12) indicates that NOEs with the TMS group are predominantly inter-residual rather than being confined to within the TMSK side chain.

DISCUSSION

The present work reports systems for genetic encoding of an unnatural amino acid with a TMS group at the end of a long, flexible side chain. A previously reported FACS-based library screening system[4] proved efficient in selecting a PylRS that recognizes TMSK. Robust amber suppression was achieved by subcloning of the PylRS/tRNA^{Pyl} pair into the high copy number pRSF plasmid with a strong T7 promoter for expression. The yield of full-length protein was enhanced further by the RF-1 knockout *E. coli* BL21(DE3) strain B95 for protein expression, enabling high cell-density fermentation without compromising the suppression fidelity. This overcomes one of the main weaknesses of PylRS/tRNA^{Pyl} pairs, namely relatively low protein yields compared to RS enzymes based on the *Mj* tyrosyl-tRNA synthetase/tRNA$_{UCA}^{Tyr}$ pair.[1] With this system, we consistently obtain proteins containing a single TMSK residue in yields no less than 50% of the yields of the corresponding wild-type proteins.

In addition to establishing a high-yielding *in vivo* system for the site-specific incorporation of TMSK based on the ChPylRS, we also successfully devised a cell-free system to produce proteins with TMSK. This was achieved by transferring the mutations from ChPylRS to the PylRS of ISO4-G1, as ChPylRS could not be overexpressed and purified in an active form. In our hands, cell-free protein synthesis with PylRS systems is practically unworkable when the N-terminal domain is present in the protein, because of the difficulty to obtain enzymatically active PylRS enzymes. ISO4-G1 contains only the C-terminal domain and its overexpression and purification in an active form is straightforward.

The ^{1}H NMR resonance of the nine protons of the TMS group in TMSK appears near 0 ppm in a spectral region with little overlap from the signals of other protein resonances. This unique feature not only facilitates the assignment of the TMS signal, especially in complex systems of high molecular weight, but also the detection of NOEs with the TMS resonance in 2D NOESY spectra, as it is much easier for automatic baseline correction algorithms to identify the baseline in spectral regions that are only sparsely populated with cross-peaks.

In addition, separation from other protein signals facilitates peak integration in 1D NMR spectra, opening a practical way to measuring the dissociation constant K_d of tightly binding ligands. Determination of small K_d values by NMR requires protein concentrations not much greater than the K_d value.[9] Depending on the line width and sensitivity of the NMR spectrometer, the intensity of the TMS signal may allow the use of submicromolar protein concentrations. Compared to ^{19}F-NMR of proteins labeled with fluorinated tags, the ^{1}H NMR signal of the TMS group is far less susceptible to line broadening by chemical shift anisotropy (CSA) relaxation and thus profits fully from NMR conducted at very high magnetic field strengths, besides offering superior sensitivity due to nine equivalent protons in a single TMS group.

In the case of the TMSK mutants of the aminoacyl-tRNA synthetase CNRS, two of the mutants displayed slow exchange on the chemical shift time scale, whereas mutation of a charged residue (Glu28) led to weaker binding and fast exchange. Unexpectedly, the mutant CNRS N117TMSK showed two TMS signals already in the absence of added amino acid, suggesting an unsymmetric dimer interface. Fitting a dissociation constant to the chemical shift changes observed for the peak maximum (Figures 6 and 7) does not fully account for the complexity of the situation in this mutant. As features of slow, intermediate, and fast exchange can be difficult to delineate if the TMS signals overlap, the best results are likely obtained by performing the experiments at the highest magnetic field strength available. Nonetheless, all fits consistently indicated that the enzyme binds the amino acid CNF more tightly than AcF. Furthermore, the titration data of all three mutants pointed to cooperativity of amino acid binding in the dimer, as indicated by significantly better fits with the Hill equation (eq S2 in the Supporting Information) than the two-state model represented by eq S1.

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

With the TMS group at the end of a long and flexible amino acid side chain, it is not surprising that its ^{1}H NMR signal is readily detected even in systems of high molecular weight such as hexameric DnaB, which has a molecular weight of over 300 kDa. Due to rapid local reorientational motions of the TMS group, its ^{1}H NMR resonance is expected to be significantly narrower than those of the protein backbone. Interestingly, however, the signal was not narrower than that observed previously for the *tert*-butyl group of *tert*-butyl tyrosine installed at the same site of DnaB.[9] At first sight, this was unexpected, as the side chain of tyrosine is considerably more rigid than the side chain of a lysine residue. It was also unexpected that the small changes in the chemical environment caused by ligand binding at fairly remote sites would be reflected by the chemical shift of the TMS resonance, which thus can serve as a probe to detect conformational changes and determine dissociation constants of protein−ligand complexes. Finally, it was unexpected, how easily NOEs could be observed with the TMS resonance. In combination, these observations indicate that the TMS group tends to interact with other hydrophobic groups on the protein surface, which limits its mobility relative to the protein and thus enhances its nuclear relaxation, but also renders its chemical shift sensitive to slight structural changes in the protein. Therefore, the length and flexibility of the tether between the TMS group and the protein backbone in the case of TMSK not only promotes reorientational motions of the TMS group relative to the protein, but also facilitates restricting these motions by allowing the TMS group to interact with hydrophobic areas of the protein which, in the present work, was evidenced by the ready observation of NOEs. In contrast, a *tert*-butyl tyrosine residue may be installed at sites, where the greater rigidity of the aromatic ring of the side chain prevents contacts of the *tert*-butyl group with other parts of the protein by projecting it firmly into the solvent, maximizing the reorientational motions of the *tert*-butyl group and minimizing its transverse relaxation rate.

Fundamentally, the unique chemical shift of the TMS resonance is a great advantage. In the present work, this allowed determination of the exchange rate between open and closed conformations of AncCDT-1, as the exchange cross-peaks are located in a region of minimal spectral overlap. Furthermore, the experiments of AncCDT-1 confirmed previous results obtained by DEER distance measurements, which had reported equal populations of open and closed conformation in the absence of ligand, and a fully closed conformation in the presence of arginine. As DEER experiments are conducted of frozen solutions in the presence of 20% glycerol, the freezing process or solvent composition could have affected the conformational equilibrium. The present experiments with TMSK indicate that this was not the case, as the intensities of the TMS resonances observed for the two states at 25 °C and in the absence of glycerol were in full agreement with the DEER results.

In the present work, we exploited the potential of the TMS group to interact with other hydrophobic groups by measuring NOEs between the TMS group and a *tert*-butyl group. Dual incorporation of a TMSK and a *tert*-butyl-tyrosine residue in response to an amber and an opal codon, respectively, was achieved by combining the ChPylRS system for TMSK with the *Mj* RS system for *tert*-butyl-tyrosine. Owing to the short-range nature of NOEs, an NOE between TMS and *tert*-butyl groups corresponds to a direct contact, delivering a structural restraint that is more precise than that obtained from paramagnetic relaxation enhancements with a nitroxide or gadolinium tag. The assignment of the NOE is straightforward, as the TMS signal is in a spectral region of the ^{1}H NMR spectrum with few other resonances and both resonances are narrow. For longer-range distance restraints, the TMS resonance can be used to detect paramagnetic relaxation enhancements generated by paramagnetic tags such as nitroxides or Gd^{3+} ions.

Strategies for making proteins with two different unnatural amino acids incorporated site-specifically in response to two different stop codons or a stop codon in combination with a quadruplet codon have been published previously,[26−28] but in our hands it has proven difficult to achieve high protein yields with these techniques. The orthogonality of PylRS enzymes with the *Mj* tyrosylRS/$tRNA_{UGA}^{Tyr}$ system has been established previously,[27] but high-yield protein expression required optimization, which we achieved by using the high copy number pRSF plasmid for PylRS expression coupled with a RF-1 knockout *E. coli* BL21(DE3) strain.

As an alternative to the incorporation of two different unnatural amino acids, it may also be possible to measure NOEs between the TMS groups of two TMSK residues, provided the interaction between the TMS groups is sufficiently long-lived to produce an NOE of significant intensity and their resonances are resolved.

In conclusion, TMSK provides a valuable probe for studying protein−protein interactions and ligand−protein binding, especially in high molecular weight macromolecules, which traditionally require isotope labeling for resonance detection and assignment. To the best of our knowledge, this is the first time that (i) PylRS mutants have successfully resulted in pure proteins with unnatural amino acids in sufficient amounts and purity for analysis by NMR spectroscopy, (ii) a PylRS system has successfully been used to produce more than 0.5 mg of protein from 1 mL of CFPS inner reaction mixture, with an unnatural amino acid other than Boc-lysine, (iii) dual incorporation of two different unnatural amino acids in the same protein was achieved in quantities sufficient for NMR spectroscopy. Although installation of unnatural amino acids potentially perturbs the properties of the target protein, it must not be overlooked that biophysical and biochemical properties can also be affected by global substitution of isotopes, as evidenced by perdeuterated proteins.[29−31] Therefore, we believe that small and flexible probes like TMSK carry outstanding potential for site-specific investigations. The plasmid pRSF-ChPylTMSK for the incorporation of TMSK is available from Addgene with the ID 163915.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.0c11971.

Synthesis protocol for TMSK, experimental procedures for protein expression and purification, NOESY pulse sequence used, FACS selection of bacterial cells with ChPylRS enzymes recognizing TMSK, mass spectrometric analysis of TMSK incorporation into PpiB and CNRS, K_d calculations from titration data, chemical structures of the C1-Gd^{3+} and MTSL tags and distributions of the distance between paramagnetic center and TMS group, sequence alignment of *Mm*PylRS, *Ma*PylRS, and G1PylRS, cell-free incorpo-

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

ration of TMSK, dual incorporation of TMSK and Tby into AncCDT-1, two-plasmid systems tested for TMSK incorporation in *E. coli* BL21(DE3), DNA sequences of the final plasmids used in the two-plasmid system, and nucleotide and corresponding amino acid sequences of the proteins used in this study (PDF)

AUTHOR INFORMATION

Corresponding Authors

Gottfried Otting − *ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia;* orcid.org/0000-0002-0563-0146; Email: gottfried.otting@anu.edu.au

Thomas Huber − *Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia;* orcid.org/0000-0002-3680-8699; Email: t.huber@anu.edu.au

Authors

Elwy H. Abdelkader − *ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia*

Haocheng Qianzhu − *Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia*

Yi Jiun Tan − *ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia*

Luke A. Adams − *ARC Training Centre for Fragment Based Design and Monash Fragment Platform, Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia;* orcid.org/0000-0002-8481-9819

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.0c11971

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. Peter G. Schultz for the pEVOL vector with the CNRS enzyme and Dr. Harpreet Vohra and Michael Devoy at the John Curtin School of Medical Research, Australian National University, for technical support on FACS experiments. Financial support by the Australian Research Council for a Laureate Fellowship to G.O. (FL170100019), project funding (DP170100162, DP200100348), and through a Centre of Excellence (CE200100012) is gratefully acknowledged.

ABBREVIATIONS

AcF = *p*-acetyl-L-phenylalanine
AMP-PNP = adenylyl-imidodiphosphate
Boc = *tert*-butoxycarbonyl
Bst = *Bacillus stearothermophilus*
ChPylRS = chimera of the N- and C-terminal domains from *Mb*RS and *Mm*RS, respectively
CFPS = cell-free protein synthesis
CNF = *p*-cyano-L-phenylalanine
CNRS = *p*-cyanophenylalanyl-tRNA synthetase
CSA = chemical shift anisotropy
DTT = dithiothreitol
EPR = electron paramagnetic resonance
FACS = fluorescence-activated cell sorting
Ma = *Methanomethylophilus alvus*
Mb = *Methanosarcina barkeri*
Mj = *Methanocaldococcus jannaschii*
Mm = *Methanosarcina mazei*
MTSL = *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1*H*-pyrrol-3-yl)methylmethanesulfonothioate
NMR = nuclear magnetic resonance
NOE = nuclear Overhauser effect
PpiB = peptidyl-prolyl *cis/trans*-isomerase B
PRE = paramagnetic relaxation enhancement
PylRS = pyrrolysyl-tRNA synthetase
RFP = red fluorescent protein
Tby = *O-tert*-butyl-tyrosine
TMS = trimethylsilyl
TMSf = *p*-TMS-phenylalanine
TMSK = N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine

REFERENCES

(1) Brabham, R.; Fascione, M. A. Pyrrolysine amber stop-codon suppression: Development and applications. *ChemBioChem* **2017**, *18*, 1973−1983.

(2) 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.

(3) Zhang, F.; Zhou, Q.; Yang, G.; An, L.; Li, F.; Wang, J. A genetically encoded 19F NMR probe for lysine acetylation. *Chem. Commun.* **2018**, *54*, 3879−3882.

(4) 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.

(5) Loh, C. T.; Adams, L. A.; Graham, B.; Otting, G. Genetically encoded amino acids with *tert*-butyl and trimethylsilyl groups for site-selective studies of proteins by NMR spectroscopy. *J. Biomol. NMR* **2018**, *71*, 287−293.

(6) Liu, Q.; He, Q.-t.; Lyu, X.; Yang, F.; Zhu, Z.-l.; Xiao, P.; Yang, Z.; Zhang, F.; Yang, Z.-y.; Wang, X.-y.; Sun, P.; Wang, Q.-w.; Qu, C.-x.; Gong, Z.; Lin, J.-y.; Xu, Z.; Song, S.-l.; Huang, S.-m.; Guo, S.-c.; Han, M.-j.; Zhu, K.-k.; Chen, X.; Kahsai, A. W.; Xiao, K.-H.; Kong, W.; Li, F.-h.; Ruan, K.; Li, Z.-j.; Yu, X.; Niu, X.-g.; Jin, C.-w.; Wang, J.; Sun, J.-p. DeSiphering receptor core-induced and ligand-dependent conformational changes in arrestin via genetic encoded trimethylsilyl ^{1}H-NMR probe. *Nat. Commun.* **2020**, *11*, 4857.

(7) Becker, W.; Adams, L. A.; Graham, B.; Wagner, G. E.; Zangger, K.; Otting, G.; Nitsche, C. Trimethylsilyl tag for probing protein-ligand interactions by NMR. *J. Biomol. NMR* **2018**, *70*, 211−218.

(8) Jabar, S.; Adams, L. A.; Wang, Y.; Aurelio, L.; Graham, B.; Otting, G. Chemical tagging with *tert*-butyl and trimethylsilyl groups for measuring intermolecular nuclear Overhauser effects in a large protein-ligand complex. *Chem. - Eur. J.* **2017**, *23*, 13033−13036.

(9) Chen, W.-N.; Kuppan, K. V.; Lee, M. D.; Jaudzems, K.; Huber, T.; Otting, G. *O-tert*-butyltyrosine, an NMR tag for high-molecular-weight systems and measurements of submicromolar ligand binding affinities. *J. Am. Chem. Soc.* **2015**, *137*, 4581−4586.

(10) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N^{ϵ}-(*o*-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **2008**, *15*, 1187−1197.

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

(11) Kaya, E.; Vrabel, M.; Deiml, C.; Prill, S.; Fluxa, V. S.; Carell, T. A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. *Angew. Chem., Int. Ed.* **2012**, *51*, 4466−4469.
(12) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **2010**, *395*, 361−374.
(13) Phillips, G. J.; Park, S.-K.; Huber, D. High copy number plasmids compatible with commonly used cloning vectors. *BioTechniques* **2000**, *28*, 400−408.
(14) van den Elzen, P. J.; Konings, R. N.; Veltkamp, E.; Nijkamp, H. J. Transcription of bacteriocinogenic plasmid CloDF13 *in vivo* and *in vitro*: structure of the cloacin immunity operon. *J. Bacteriol.* **1980**, *144*, 579−591.
(15) Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K. Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* **2015**, *5*, 9699.
(16) 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.
(17) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. A genetically encoded infrared probe. *J. Am. Chem. Soc.* **2006**, *128*, 13984−13985.
(18) Wilding, M.; Hong, N.; Spence, M.; Buckle, A. M.; Jackson, C. J. Protein engineering: the potential of remote mutations. *Biochem. Soc. Trans.* **2019**, *47*, 701−711.
(19) Yagi, H.; Banerjee, D.; Graham, B.; Huber, T.; Goldfarb, D.; Otting, G. Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. *J. Am. Chem. Soc.* **2011**, *133*, 10418−10421.
(20) Battiste, J. L.; Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* **2000**, *39*, 5355−5365.
(21) Willis, J. C. W.; Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* **2018**, *10*, 831−837.
(22) Seki, E.; Yanagisawa, T.; Kuratani, M.; Sakamoto, K.; Yokoyama, S. Fully productive cell-free genetic code expansion by structure-based engineering of *Methanomethylophilus alvus* pyrrolysyl-tRNA synthetase. *ACS Synth. Biol.* **2020**, *9*, 718−732.
(23) Loscha, K. V.; Herlt, A. J.; Qi, R.; Huber, T.; Ozawa, K.; Otting, G. Multiple-site labeling of proteins with unnatural amino acids. *Angew. Chem., Int. Ed.* **2012**, *51*, 2243−2246.
(24) Clifton, B. E.; Kaczmarski, J. A.; Carr, P. D.; Gerth, M. L.; Tokuriki, N.; Jackson, C. J. Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* **2018**, *14*, 542−547.
(25) Kaczmarski, J. A.; Mahawaththa, M. C.; Feintuch, A.; Clifton, B. E.; Adams, L. A.; Goldfarb, D.; Otting, G.; Jackson, C. J. Altered conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme. *Nat. Commun.* **2020**, *11*, 5945.
(26) Kim, J.; Seo, M.-H.; Lee, S.; Cho, K.; Yang, A.; Woo, K.; Kim, H.-S.; Park, H.-S. Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis. *Anal. Chem.* **2013**, *85*, 1468−1474.
(27) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **2013**, *52*, 1828−1837.
(28) Lammers, C.; Hahn, L. E.; Neumann, H. Optimized plasmid systems for the incorporation of multiple different unnatural amino acids by evolved orthogonal ribosomes. *ChemBioChem* **2014**, *15*, 1800−1804.
(29) Hattori, A.; Crespi, H. L.; Katz, J. J. Effect of side-chain deuteration on protein stability. *Biochemistry* **1965**, *4*, 1213−1225.
(30) Piszczek, G.; Lee, J. C.; Tjandra, N.; Lee, C.-R.; Seok, Y.-J.; Levine, R. L.; Peterkofsky, A. Deuteration of *Escherichia coli* enzyme I^{Ntr} alters its stability. *Arch. Biochem. Biophys.* **2011**, *507*, 332−342.
(31) Nichols, P. J.; Falconer, I.; Griffin, A.; Mant, C.; Hodges, R.; McKnight, C. J.; Vögeli, B.; Vugmeyster, L. Deuteration of nonexchangeable protons on proteins affects their thermal stability, side-chain dynamics, and hydrophobicity. *Protein Sci.* **2020**, *29*, 1641−1654.

https://dx.doi.org/10.1021/jacs.0c11971
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX