Probing Ligand Binding Sites on Large Proteins by NMR Spectroscopy of Genetically Encoded Non-Canonical Amino Acids

Kasuni B. Ekanayake,[†] Mithun C. Mahawaththa,[†] Haocheng Qianzhu,[‡] Elwy H. Abdelkader,[†] Josemon George,[‡] Sven Ullrich,[‡] Rhys B. Murphy,[‡] Sarah E. Fry,[§+] Jason Johansen-Leete,[§+] Richard J. Payne,[§+] Christoph Nitsche,[‡] Thomas Huber,[‡] and Gottfried Otting[†,*]

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

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

[§] Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, The University of Sydney, Sydney, NSW 2006, Australia

[+] School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia

ABSTRACT

N^6-(((trimethylsilyl)-methoxy)carbonyl)-L-lysine (TMSK) and N^6-trifluoroacetyl-L-lysine (TFAK) are non-canonical amino acids, which can be installed in proteins by genetic encoding. In addition, we describe a new aminoacyl-tRNA synthetase specific for N^6-(((trimethylsilyl)methyl)-carbamoyl)-L-lysine (TMSNK), which is chemically more stable than TMSK. Using the dimeric SARS-CoV-2 main protease (M^{pro}) as a model system with three different ligands, we show that the ^{1}H and ^{19}F nuclei of the solvent-exposed trimethylsilyl and CF_3 groups produce intense signals in the nuclear magnetic resonance (NMR) spectrum. Their response to active-site ligands differed significantly when positioned near rather than far from the active site. Conversely, the NMR probes failed to confirm the previously reported binding site of the ligand pelitinib, which was found to enhance the activity of M^{pro} by promoting the formation of the enzymatically active dimer. In summary, the amino acids TMSK, TMSNK, and TFAK open an attractive path for site-specific NMR analysis of ligand binding to large proteins of limited stability and at low concentration.

INTRODUCTION

The detection of ligand binding sites on proteins can be challenging, if the system is amenable to neither protein crystallography nor NMR analysis of isotope-labelled samples. The situation is particularly difficult if the effect of ligand binding cannot be probed by an enzymatic assay. This case is commonly addressed by systematic site-directed mutagenesis experiments,[1–3] but some mutations may not affect ligand binding while others can affect ligand binding via allosteric effects. The present work explores the use of three different genetically encoded non-canonical amino acids for the detection of ligand binding by NMR spectroscopy. The amino acids carry an NMR probe at the end of a long and flexible tether. In this way, they minimize the impact from allosteric effects mediated through the backbone of the protein, while allowing exploration of a greater radius around the probe to more easily sense the presence of a ligand in its vicinity. The much greater mobility of the probe relative to the protein backbone adds the advantage of ready detection by NMR spectroscopy even when installed in a large protein that is not isotope labelled. In addition, they yield high intensity NMR signals that can be detected at low protein concentration. The present work used the SARS CoV-2 main protease (M^{pro}) to benchmark the NMR detection of ligand binding using these NMR probes. M^{pro} forms a homodimer of about 67 kDa molecular weight, that is difficult to analyze by conventional NMR spectroscopy and displays pronounced allosteric effects upon the binding of inhibitors to the active site.[4]

Figure 1 shows the chemical structures of the non-canonical amino acids used in the present work, for which orthogonal systems for genetic encoding have recently become available.[5–7] The amino acid N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine (TMSK; Figure 1a) exploits the ^{1}H NMR chemical shift of the trimethylsilyl (TMS) group, which is near 0 ppm and thus in a spectral region where there is little overlap from the signals of any biological or buffer molecules. When positioned at a highly solvent-exposed site, the nine equivalent ^{1}H spins in the TMS group yield an intense and narrow NMR signal that is more akin to that of a low molecular weight compound than the protein it is attached to, enabling its detection in proteins of > 300 kDa molecular weight. Nonetheless, the exact chemical shift and line width is sensitive to its chemical environment and can serve as an indicator of ligand binding.[6]

The amino acid N^6-trifluoroacetyl-L-lysine (TFAK, Figure 1c) is similar to TMSK but designed for detection of the CF_3 group by ^{19}F NMR spectroscopy. As the canonical amino acids do not contain fluorine and ^{19}F occurs with 100% natural abundance, the ^{19}F NMR signal of TFAK can be observed free from spectral overlap and with good sensitivity. ^{19}F spins feature a much greater range of chemical shifts than ^{1}H spins,[8] but ^{19}F relaxation tends to be faster than ^{1}H relaxation at high magnetic field strengths due to large chemical shift anisotropies, and a CF_3 group contains only three equivalent spins compared to nine in a TMS group. Nonetheless, probes containing trifluoromethyl groups have become popular reagents for site-specific monitoring of protein responses to ligand binding by ^{19}F-NMR spectroscopy.[9,10]

Some of the M^{pro} samples prepared with TMSK proved to be sensitive to degradation. We therefore aimed to improve the chemical stability of the TMS probe by synthesizing N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine (TMSNK; Figure 1b), where the carbamate group of TMSK is replaced by a urea functionality. To enable site-specific incorporation in response to an amber stop codon like for TMSK and TFAK, we selected a new aminoacyl-tRNA synthetase for this amino acid. This allowed comparing the NMR performance of TMSK, TFAK, and TMSNK.

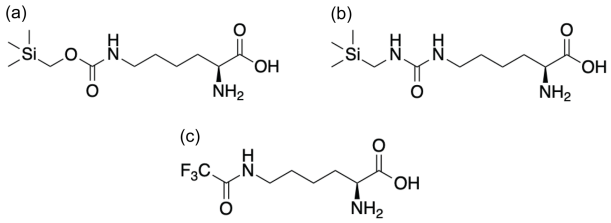

Figure 1. Chemical structures of the non-canonical amino acids used. (a) N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine (TMSK). (b) N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine (TMSNK) (c) N^6-trifluoroacetyl-L-lysine (TFAK).

From the numerous ligands that have been reported for the SARS-CoV-2 M^{pro} enzyme[11,12] we selected three to benchmark the performance of different non-canonical amino acids. The cyclic peptide inhibitor **1** (Figure 2a) had previously been identified by us as a tight binding active-site inhibitor, but the presence of inhibitor **1** elicited extensive allosteric effects, which made it impossible to map its binding site by chemical shift

changes of backbone amides.[4] Calpeptin (Figure 2b) and pelitinib (Figure 2c) are small M^{pro} ligands, for which crystal structures in complex with M^{pro} have been reported.[12]

Using the M^{pro}–peptide complexes as model systems, the present study benchmarks the use of TMSK, TFAK, and TMSNK as genetically encoded NMR probes of ligand binding by monitoring their responses to ligand binding following installation at different solvent-exposed sites. In addition, we checked the effect of the mutations on the enzymatic activity in a functional assay and compared the activity of TMSK mutants with a more traditional mutagenesis approach, where selected sites are mutated to alanine.

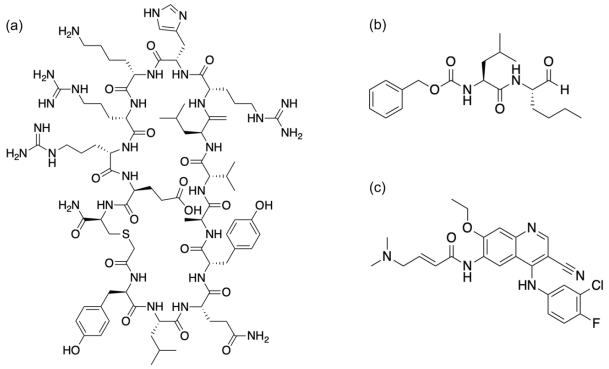

Figure 2. Chemical structures of ligands used. (a) Cyclic peptide inhibitor **1**. (b) Calpeptin. (c) Pelitinib.

RESULTS

Site-Specific Incorporation of Non-Canonical Amino Acids. Based on the co-crystal structures of SARS-CoV-2 M^{pro} complexed with ligands (PDB ID 7RNW for inhibitor **1**,[4] 7AKU for calpeptin, and 7AXM for pelitinib,[12] ten solvent-exposed residues of M^{pro} were targeted for mutation to either TMSK, TMSNK, TFAK, or alanine (Figure 3). As the aminoacyl-tRNA synthetase for TMSK did not recognize TMSNK, a new synthetase was selected for this amino acid as described in the Experimental Section. TMSK and TFAK residues were successfully introduced at the selected locations using previously reported genetic encoding systems.[6,7] Mass spectrometry indicated excellent amber suppression efficiency (Figure S10–S12).

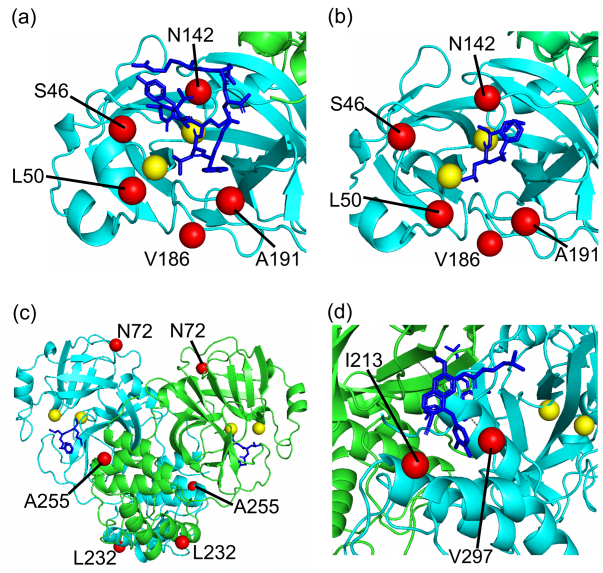

Figure 3. Sites of M^{pro} targeted for mutation to TMSK, TMSNK, TFAK, or alanine. The locations of the mutated amino acid residues are identified by red balls plotted on a ribbon representation of the co-crystal structures of M^{pro} with the different ligands, showing the individual M^{pro} monomers in green and cyan, respectively, and the bound ligands in a blue stick representation. Yellow spheres identify the active site residues His41 and Cys145. (a) Sites selected close to the binding site of inhibitor **1** (PDB ID: 7RNW[4]). The crystal structure does not report electron density for all residues of the inhibitor **1**. (b) Same as (a), but for the complex with calpeptin (PDB ID: 7AKU[12]). (c) Mutation sites far from the active site plotted on the calpeptin–M^{pro} structure. (d) Location of the allosteric ligand pelitinib relative to two of the mutation sites targeted in this work (PDB ID: 7AXM[12]).

Titrations of TMSK mutants with active-site and allosteric ligands. The ^{1}H NMR spectra recorded of the purified protein samples of TMSK and TMSNK mutants showed clear and well-resolved signals for the TMS group at chemical shifts near 0 ppm, except for the mutant A191TMSK, which showed very broad signals at two different chemical shifts, suggesting line broadening by an exchange between different conformations (Figure 4e). The intensity and unique chemical shift of the TMS resonance near 0 ppm made its assignment straightforward. The line widths of the TMS signals varied between different mutation sites but could be as narrow as 10 Hz.

In titration experiments conducted with increasing concentrations of ligands, the NMR signals of TMSK at different sites responded in fundamentally different ways depending on the distance of the TMSK residue from the inhibitor binding site. S46, L50, N142, V186,

and A191 are all in close proximity of inhibitor **1**, and a TMS group at the end of the side chain of TMSK could plausibly make direct contacts with the ligand. For the mutants S46TMSK and V186TMSK, the titration experiments revealed a new signal for the complex while the original TMS signal disappeared (Figure 4a and b). The signals of the free and bound protein could be discerned even when separated by as little as 0.02 ppm, and their coexistence at intermediate titration ratios indicated slow exchange of the inhibitor between different protein molecules on the low millisecond timescale. For the mutants L50TMSK and N142TMSK, the signal observed for the apo-protein disappeared upon addition of inhibitor **1**. The disappearance of the TMS signal of the N142TMSK mutant can be explained either by exchange broadening in the complex or immobilization of the TMS group, e.g., by being trapped at the protein–ligand interface. The crystal structure shows N142 in the center of the inhibitor binding site (Figure 3a).

In the titration experiment of the mutant L50TMSK, the TMS signal disappeared right after the addition of 0.1 equivalents of inhibitor (Figure 4d), indicating rapid exchange between free and inhibitor-bound protein. This mutant thus appears to bind the inhibitor with faster kinetics than the N142TMSK mutant. In contrast, titration of the mutant A191TMSK yielded a narrower TMS signal with inhibitor than for the free protein, which suggests exchange-broadening of the signal in the apo-protein and transition to a single conformational species in the complex (Figure 4e). In summary, the TMS signal of residues near the ligand binding site responded to the presence of inhibitor by changing its chemical shift or drastically changing its line width, including broadening beyond detection or appearance of the signal upon formation of the protein–ligand complex. In contrast, TMSK residues positioned at sites far from the ligand binding site were far less sensitive to the addition of inhibitor **1**, as illustrated by the high conservation of the TMS signals of the mutants N72TMSK, L232TMSK, and A255TMSK (Figure 4f–h). The most pronounced spectral change for these mutants is some line broadening or weakening of the signals observed for the 1:1 complexes, which may reflect altered protein dynamics.

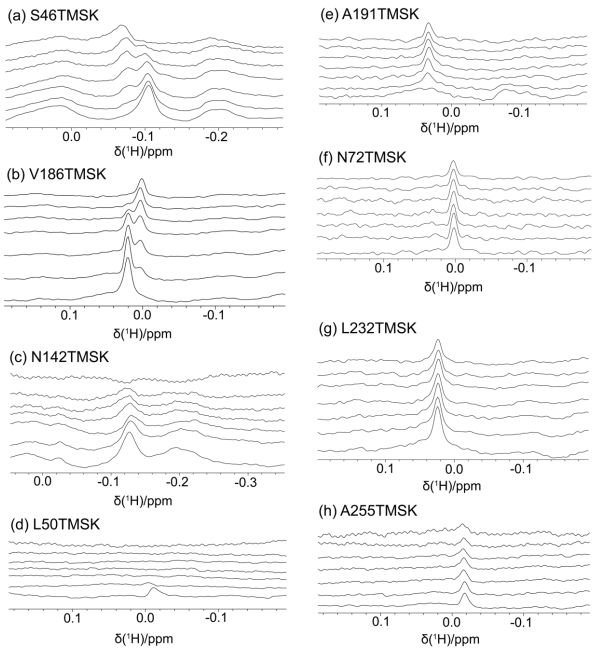

Figure 4. Selected spectral regions of 1D ^{1}H NMR spectra showing the signal of the TMS group of M^{pro} variants with a TMSK residue positioned at different sites. The mutation sites are indicated in the subfigures. The spectra were recorded of 30 μM solutions of M^{pro} in 90% H_2O/10% D_2O titrated with a 1 mM stock solution of inhibitor **1**. The titration ratio of inhibitor to protein was, from bottom to top, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5 and 1.0. (a) – (e) Mutation sites close to the active site. (f) – (h) Mutation sites remote from the active site.

As inhibitor **1** occupies an extended site on the protease, we repeated the titrations of the TMSK mutants with the much smaller active-site inhibitor calpeptin (Figures 2b and 3b). As expected for an inhibitor with a smaller footprint, the spectral changes observed for TMSK in positions 46, 186, 142, 50, and 191 were smaller but nonetheless significant. Titrations of the S46TMSK and N142TMSK mutants led to the disappearance of the TMS signal (Figure 5a and c), while the mutants V186TMSK and L50TMSK yielded a new, albeit poorly resolved signal (Figure 5b and d), and the mutant A191TMSK produced a TMS signal in the complex that was absent in the apo-protein (Figure 5e). As in the titration experiment with inhibitor **1**, the TMS signal of the mutants remote from the active site did not change their chemical shift and any changes in line width were much less pronounced (Figure 5f–h). In summary, the effects generated by inhibitor **1** were broadly recapitulated by calpeptin despite its smaller size.

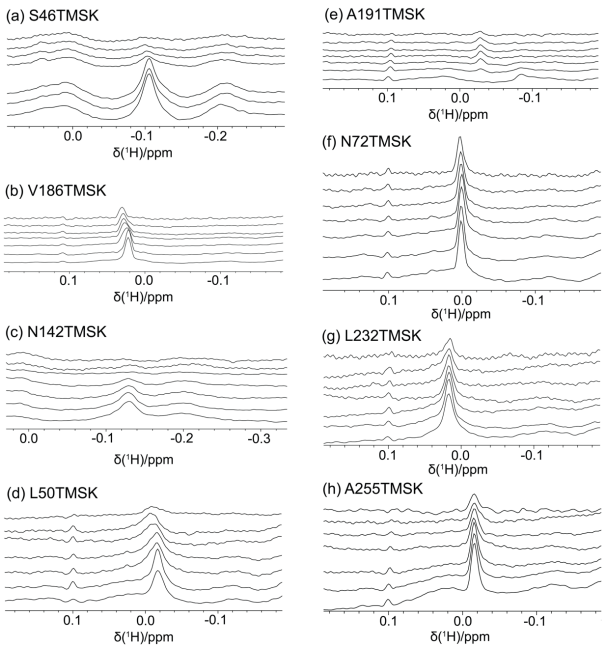

Figure 5. Selected spectral regions of 1D ^{1}H NMR spectra of TMSK mutants of M^{pro} titrated with calpeptin instead of inhibitor **1**. All other parameters were the same as in Figure 4. (a) – (e) Mutation sites close to the active site. (f) – (h) Mutation sites remote from the active site.

To compare the effects of an allosterically binding compound with those produced by the active-site inhibitors, we also titrated the various TMSK mutants with the allosteric ligand pelitinib (Figure 2c). In addition, we prepared the mutants I213TMSK and V297TMSK to probe sites in close proximity of the pelitinib binding site determined by crystallography (Figure 3d). Figure 6 shows that all sites probed retained the ^{1}H chemical shifts of the TMS groups, and the main change observed was limited to a slight increase in the line width at greater ligand-to-protein titration ratios. The spectral changes elicited by the allosteric ligand are thus distinctly different from those generated by the active-site inhibitors. In addition, there is no evidence that pelitinib binds near I213 and V297.

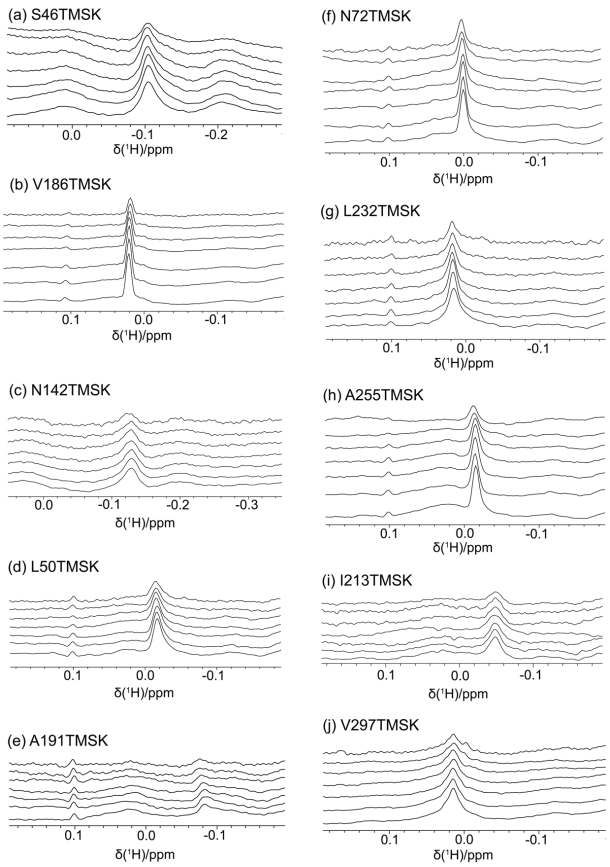

Figure 6. Selected spectral regions of 1D ^{1}H NMR spectra of TMSK mutants of M^{pro} titrated with pelitinib. All other parameters were the same as in Figures 3 and 4. (a) – (h) Mutation sites far from the crystallographically determined binding site.[12] (i) and (j) Mutation sites close to the crystallographic binding site.

Titrations of TFAK Mutants. Next, we compared the effects displayed by TFAK in the ^{19}F NMR spectrum with those displayed by TMSK in the ^{1}H NMR spectrum. The ^{19}F NMR spectra of different TFAK mutants showed clear signals for the CF_3 group with chemical shifts varying by up to 1 ppm and line widths as narrow as 24 Hz. Trifluoroacetic acid (TFA) is routinely used to deprotect chemically synthesized peptides. The ^{19}F NMR signal of TFA provides a welcome reference for calibrating the ^{19}F chemical shifts[8] but also interferes with the signal from the CF_3 group of TFAK. In the case of TFAK in positions 46, 186, 191, and 72, the titration with inhibitor **1** led to the appearance of a new signal that was detectable despite the dominating presence of TFA (Figure 7). A notable high-field shift of the CF_3 signal of TFAK at the maximum titration ratio of 1:1 versus the 0.5:1 titration ratio arose from the DMSO present in the stock solution of the inhibitor and could be reproduced by titration with DMSO alone (Figure S13). The sensitivity to solvent

highlights the significantly greater sensitivity of the ^{19}F chemical shift of TFAK towards small changes in the chemical environment compared with the ^{1}H chemical shift of TMSK. This greater sensitivity is also reflected by the appearance of a new signal for the N72TFAK mutant in the M^{pro}–inhibitor complex (Figure 7f) despite N72 being remote from the active site (Figure 3c). To verify the findings with less interference from the TFA signal, we repeated the titrations of V186TFAK, N142TFAK and L50TFAK with inhibitor **1**, which had been HPLC-purified with acetonitrile containing formic acid instead of TFA. While the inhibitor sample unfortunately still contained traces of TFA, the results fully agreed with the data of Figure 7, suggesting signal overlap between TFAK and TFA solvent (Figure S14).

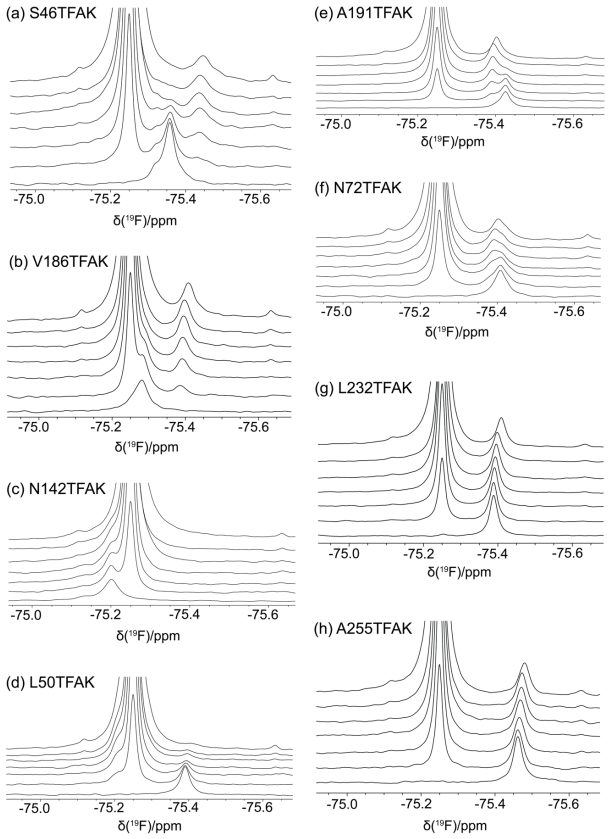

Figure 7. Selected spectral regions of 1D ^{19}F NMR spectra of TFAK mutants of M^{pro} titrated with inhibitor **1**. The titration experiments were conducted with 0.1 mM solutions of M^{pro} with 2.5 mM stock solution of inhibitor. The same titration ratios were explored as in Figures 4–6. Control experiments showed that the upfield change in chemical shift observed between the titration ratios 0.5:1 and 1:1 (top two traces in each panel) arises from the DMSO solvent of the inhibitor stock solution (Figure S13). Trifluoroacetate in the preparation of inhibitor **1** contributes to the large signal intensity at -75.25 ppm. (a) – (e)

Mutation sites close to the active site. (f) – (h) Mutation sites remote from the active site. See Figure S14 for titration data of the mutants L50TFAK, N142TFAK, and V186TFAK, using inhibitor **1** HPLC-purified with formic acid instead of trifluoroacetic acid additive in the running solvent.

To probe the performance of TFAK with a different active-site inhibitor, we repeated the titrations with calpeptin (Figure 8). The results discriminated very clearly between mutations close to and far from the inhibitor binding site, in that all TFAK mutants featured a new chemical shift for the M^{pro}–inhibitor complex when they were close to the inhibitor binding site (Figure 8a–e) but unchanged chemical shift when they were located far from the active site (Figure 8f–h). This response pattern is the same as detected with inhibitor **1** (Figure 7), except that N72TFAK, which is a mutant far from the active site, resolved a new chemical shift for the complex with inhibitor **1** (Figure 7f), indicating an allosteric response.

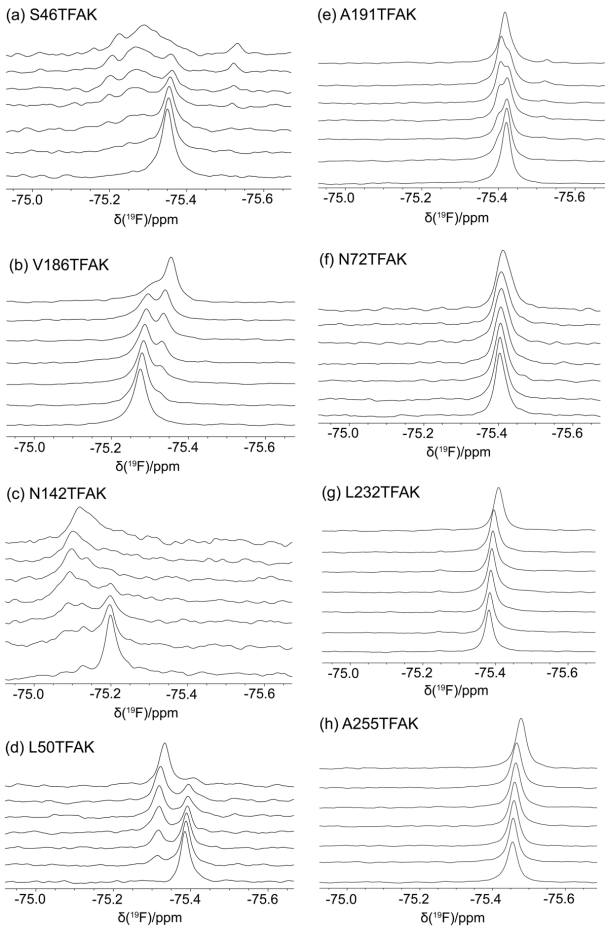

Figure 8. Selected spectral regions of 1D ^{19}F NMR spectra showing the TFAK resonance of different M^{pro} variants as a function of added calpeptin. (a) – (e) Mutation sites close to the active site. (f) – (h) Mutation sites remote from the active site. All other parameters were the same as in Figure 7.

To further explore the potential of TFAK mutants for allosteric responses, we repeated the titrations with the allosteric ligand pelitinib (Figure 3d). Fully consistent with the results obtained with TMSK, none of the TFAK mutants, including the mutants I213TFAK and V297TFAK, displayed a new signal for the complex with pelitinib (Figure 9). The ^{19}F NMR signal of TFAK at sites 213 and 297 was significantly broader than at any of the other sites, suggesting less flexibility of the TFAK sidechain due to limited solvent exposure. Notably, the ^{1}H NMR signal of TMSK at these sites is also relatively broad (Figure 6i and j), but the effect is much smaller than for TFAK, in accord with the longer sidechain of TMSK.

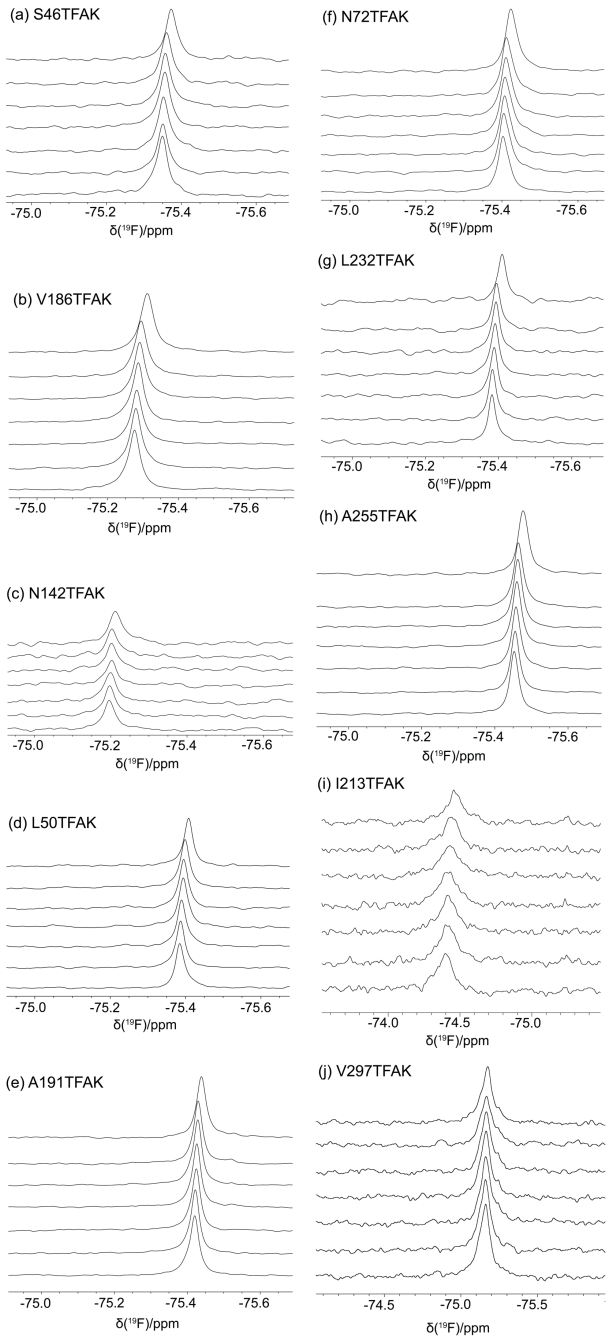

Figure 9. Selected spectral regions of 1D ^{19}F NMR spectra of TFAK mutants of M^{pro} titrated with pelitinib. (a) – (e) Mutation sites close to the active site. (f) – (j) Mutation sites remote from the active site. All other parameters were the same as in Figures 7 and 8.

TMSNK, a New NMR Probe of Ligand Binding. Some of the TMSK mutants proved to degrade during storage at -20 ^{o}C, resulting in the gradual loss of the ^{1}H NMR signal of the TMS group. When this had been observed for the mutants N72TMSK and V186TMSK, we switched to storage at -80 ^{o}C, which remediated the problem. To address the possibility

that the carbamate group of TMSK is prone to hydrolysis, we synthesized the urea analogue TMSNK (Figure 1b), which was presumed to be more stable. As the aminoacyl-tRNA synthetase used for genetic encoding of TMSK did not recognize TMSNK, a new aminoacyl-tRNA synthetase was selected from a library of pyrrolysyl-tRNA synthetases derived from the methanogenic archaeon ISO4-G1 (G1PylRS).[6] With the new aminoacyl-tRNA synthetase at hand, we tested its performance for calpeptin and pelitinib complexes of N72TMSNK. Compared with the N72TMSK mutant, the chemical shift of the TMS group was shifted high-field by 0.01 ppm. As expected for the location of residue 72 far from the active site, the titrations with the ligands caused no change in chemical shift (Figure 10a and c). Like the titrations of N142TFAK with calpeptin (Figure 8c), the titration of N142TMSNK with calpeptin indicated conformational heterogeneity in the complex with the inhibitor (Figure 10b) in agreement with the very close location of site 142 to the inhibitor (Figure 3b).

Like its progenitor TMSK, the new amino acid TMSNK thus proved a sensitive reporter of ligand binding. However, while good protein yields were obtained for TMSNK mutants at sites N72 and N142, TMSNK mutants of sites V186 and A191 yielded too little protein for NMR studies. To test whether TMSNK mutants are more stable than TMSK mutants, N72TMSK and N72TMSNK mutants were produced in parallel. Aliquots were frozen and stored at –20 °C. When the ^{1}H NMR signal of the TMS group was recorded after one week, the TMSK mutant was mostly degraded, whereas the ^{1}H NMR signal of the TMS group of TMSNK mutant remained mostly intact (Figure S15).

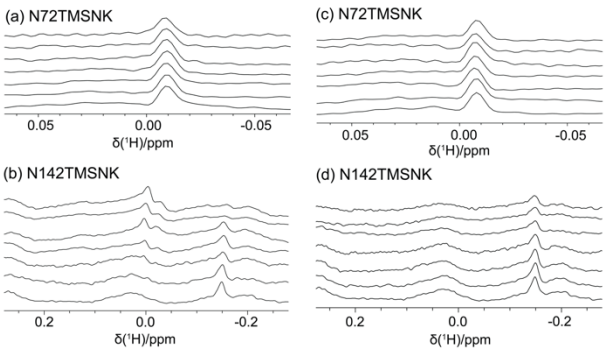

Figure 10. Selected spectral regions of 1D ^{1}H NMR spectra of TMSNK mutants of M^{pro} titrated with calpeptin (a and b) and pelitinib (c and d). The sample conditions were the same as in Figure 4.

Enzyme Activity. Active-site inhibitors such as the inhibitor **1** affect the activity of wild-type M^{pro} in a concentration dependent manner.[4] The dissociation constant of dimeric wild-type M^{pro} has been reported as 7 mM (measured by small angle X-ray scattering),[13] 2.5 mM (measured by ultracentrifugation),[14] and 0.14 mM (measured by mass spectrometry).[15] At the concentration used for the enzymatic activity assay (25 nM), a significant fraction of the enzyme thus was in the monomeric state, complicating the data analysis as the activity of monomeric M^{pro} is known to be negligible or absent.[14,16] Figure 11 shows the activity data normalized to the activity of the respective mutants in the experiments conducted without ligand.

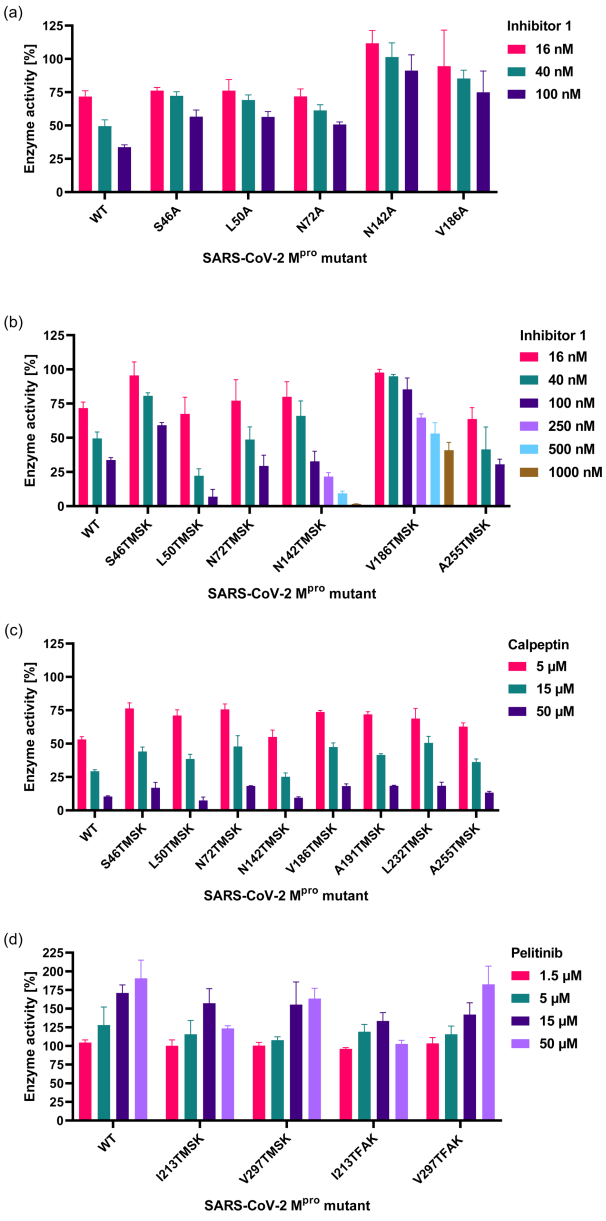

Figure 11. Sensitivity of the enzymatic activity of wild-type (WT) and mutant M^{pro} towards inhibitor **1**, calpeptin, and pelitinib. The enzymatic activities are given in percent of the activity observed without ligand. (a) WT and different alanine mutants titrated with inhibitor **1**. (b) WT and different TMSK mutants titrated with inhibitor **1**. (c) WT and different TMSK mutants titrated with calpeptin. (d) WT and different TMSK and TFAK mutants titrated with pelitinib.

All TMSK mutants displayed enzymatic activity, demonstrating that structural integrity was maintained in the variants. Furthermore, their activity decreased with increasing concentrations of inhibitor **1** or calpeptin, confirming inhibitor binding (Figure 12b and c). For comparison with more standard mutagenesis experiments, we also produced alanine mutants of all the sites investigated with TMSK and TFAK. All alanine mutants proved enzymatically active, and their activity was suppressed by the inhibitor **1** in a concentration dependent manner (Figure 11a). For the same concentrations of inhibitor **1** or calpeptin, the activities of most mutants were comparable to that of the wild type, but the mutants S46TMSK, V186TMSK, N142A, and V186A displayed a lesser sensitivity towards the inhibitor **1** than the wild type, and the mutant L50TMSK showed a greater sensitivity towards the inhibitor **1**. The magnitude of these effects did not correlate with the distance of the mutation sites from the inhibitor binding site. Most strikingly, the mutant N142TMSK maintained its enzymatic activity, which was lowered by inhibitor **1** with a potency comparable to the wild type (Figure 11b), even though this site is located right at the inhibitor binding site, and the mutant N142L has been shown to raise the IC_{50} value of inhibitor **1** 5-fold relative to the wild type.[17]

Unexpectedly, we found no evidence that pelitinib inhibits wild-type M^{pro} (Figure 12d). To the contrary, the enzymatic activity was enhanced by increasing concentrations of pelitinib, and the mutants V297TMSK and V297TFAK showed the same effect. The mutants I213TMSK and I213TFAK showed a maximum in activity at 15 mM pelitinib, suggesting some inhibitory effect only at much greater concentration of pelitinib. Due to the poor solubility of pelitinib in aqueous solution, no experiments were conducted at pelitinib concentrations greater than 50 mM.

Dimerization of M^{pro} Promoted by Pelitinib. To explore whether pelitinib enhances the activity of M^{pro} by stabilizing the enzymatically active dimer, we combined size-

exclusion chromatography (SEC) experiments with molecular weight determination by multiangle light scattering (MALS). SEC-MALS experiments conducted with 1 mM solutions of M^{pro} indicated an effective mass between that of the monomer (35 kDa for our His_6-tagged construct) and dimer (70 kDa) as expected for an equilibrium between both states (Figure 12a). The molecular weight of the dimer was detected in the presence of a two-fold excess of pelitinib (Figure 12b). Calpeptin also shifted the equilibrium towards the dimer, but less strongly (Figure 12c and d). This suggests that the increase in enzymatic activity arises from pelitinib changing the monomer-dimer equilibrium towards the active dimer of M^{pro}.

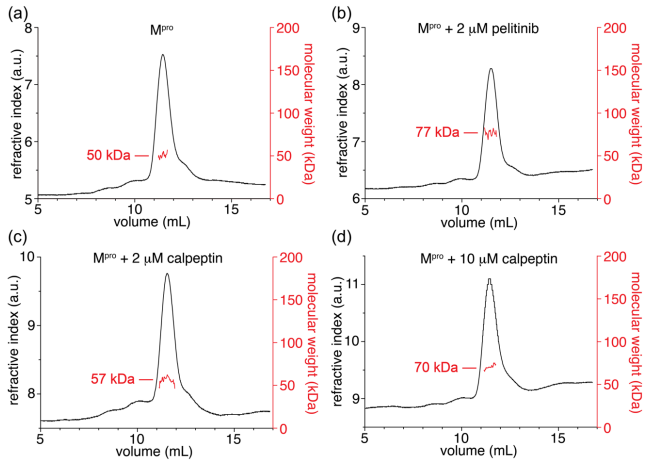

Figure 12. SEC-MALS of M^{pro} indicates dimerization in the presence of pelitinib or calpeptin. The experiments were conducted with 1 mM M^{pro} in a buffer of 20 mM Tris–HCl, pH 8.0, 150 mM NaCl. The refractive index and molecular weight determined by MALS are plotted in black and red, respectively, versus the elution volume of the size exclusion chromatography. Experiments were conducted with free protein (a) and in the presence of 2 mM pelitinib (b), 2 mM calpeptin (c), or 10 mM calpeptin (d).

DISCUSSION

Non-canonical amino acid probes for ligand detection by NMR. This work demonstrates the utility of the recently reported genetically encoded amino acids TMSK[6] and TFAK[7] as site-specific NMR probes for the detection of ligand binding, and the new amino acid TMSNK is shown to perform as a probe with properties closely related to TMSK. Every mutant targeted by installing single amber stop codons in the gene of M^{pro} was successfully produced. Despite the high molecular weight of the M^{pro} dimer (about 70

kDa), the ^{1}H NMR and ^{19}F NMR signals of the TMS and CF_3 groups, respectively, were readily observed without isotope labelling and despite relatively low protein concentrations (30 mM for TMSK and TMSNK mutants, 100 mM for TFAK mutants, using 3 mm NMR tubes). The homodimer of the SARS-CoV-2 main protease is a challenging system, as it digests the cyclic peptide inhibitor **1**, limiting the time for measurements of the intact complex. Owing to the high intensity of the TMS resonance, which presents a singlet produced by nine equivalent protons, the ^{1}H NMR measurement times could be limited to 5 minutes per spectrum.

All three probes proved sensitive reporters of the nearby presence of large and small ligands, displaying either new signals for the protein–ligand complexes or significant line broadening or line narrowing effects, whereas the spectral changes were much less pronounced when the probes were installed in locations remote from the ligand binding site. In previous work with isotope-labelled wild-type M^{pro}, we observed significant chemical shift perturbations and line width effects for the amides of L232 and A255, which are far from the binding site of the inhibitor **1** and therefore could only be explained by allosteric effects.[4] This compromised the attempt to identify the ligand binding site by chemical shift perturbations. The probes of the present work proved far more resilient towards allosteric effects. Only the mutant N72TFAK displayed a new chemical shift in the complex with inhibitor **1** despite being remote from the inhibitor binding site, but the change in chemical shift was small and not reproduced by the complex with the smaller inhibitor calpeptin (Figures 7f and 8f). The lesser sensitivity of the non-canonical amino acid probes towards allosteric effects compared with backbone amides may stem from the flexibility of the lysine side chains linking the probes to the polypeptide backbone, which dampens any transmission of backbone conformational changes via the amino acid sidechain to the probe. In contrast, if the probes make van der Waals contacts with a bound ligand, the chemical environment of the TMS and CF_3 groups can change substantially, especially if the ligand allows new hydrophobic interactions, as both groups are hydrophobic moieties. Importantly, for probes located at or near the ligand binding site, no TMS or CF_3 signal remained unchanged in the titration experiments, indicating a low probability of false negative results.

The ^{19}F NMR signal of the CF_3 group in TFAK proved to be superior to the ^{1}H NMR signal of the TMS group in TMSK or TMSNK, both with regard to simpler peak

identification (owing to the absence of background signal from any of the other amino acids) and greater changes in chemical shifts (owing to the intrinsically greater sensitivity of ^{19}F chemical shifts to different chemical environments). Drawbacks associated with TFAK are the lesser sensitivity of detection, which we compensated by using higher protein concentrations, the common presence of TFA solvent in synthetic peptide samples, and the somewhat greater line width resulting from the important contribution of chemical shift anisotropy effects to nuclear ^{19}F relaxation, which may render small changes in chemical shifts more difficult to resolve. On a practical note, our aminoacyl-tRNA synthetase for TFAK is less efficient than the corresponding enzymes for TMSK and TMSNK. Protein mutants with TFAK were still produced in good yield with the provision of TFAK in greater concentration (20 mM versus 1 mM TMSK and TMSNK). TFAK is readily available commercially and inexpensive. The synthetic route to TMSNK is straightforward (Figure 13).

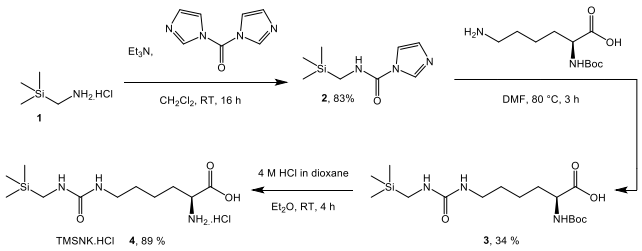

Figure 13. Synthetic scheme of TMSNK hydrochloride.

Properties of the non-canonical amino acid NMR probes TMSK, TMSNK, and TFAK. The amino acids TMSK, TMSNK, and TFAK are all based on lysine, which features a long and flexible side chain. With the TMS and CF_3 probes positioned at the end of the side chain, they reorientate much more quickly relative to the NMR magnet than the backbone of the protein they are attached to. The resulting short rotational correlation times of the probes are the basis for producing narrow NMR signals. Despite their enhanced flexibility, the TMS and CF_3 groups sense the local chemical environment by transient contacts, which do not necessarily broaden the NMR resonances very much. These features have previously been observed also for TMS and *tert*-butyl compounds chemically ligated to cysteine residues in systems as large as 95 kDa.[18] As M^{pro} contains twelve cysteine residues, including the active site residue Cys145, however, chemical modification of single cysteine residues is not feasible for this enzyme without extensive mutations of natural cysteine residues. In a similar vein, mutation of natural cysteine residues coupled

with chemical ligation of CF_3 probes to single cysteine residues positioned at strategically selected positions has been deployed extensively to monitor the effect of different ligands binding to G-protein coupled receptors (GPCR) by NMR spectroscopy.[19] Only recently a genetically encoded amino acid bearing a CF_3 group, *m*-CF_3-phenylalanine, has been introduced into a GPCR for detection by NMR.[20] The availability of a genetic encoding system, which allows the facile incorporation of an NMR probe at a single site, presents an important advance, but enhanced flexibility of the NMR probe with respect to the core of the protein remains critically important to limit the line widths observed in high molecular weight systems. For example, the genetic encoding systems recently established for *p*-SF_5-phenylalanine and 7-fluoro-tryptophan delivered ^{19}F NMR line widths approaching 100 Hz and more for relatively small proteins of molecular weights below 26 kDa.[21,22]

Owing to the flexibility of their side chains, the amino acid probes of the present work compare very favorably with other, less flexible, non-canonical amino acids containing a TMS group. Furthermore, the ^{1}H chemical shift of the TMS group is in a more favorable spectral region when it is linked to a non-aromatic moiety. For example, the chemical shift of the TMS group in *p*-(trimethylsilyl)phenylalanine is about 0.2 ppm more low-field than in TMSK and thus more prone to spectral overlap with protein signals,[23] and the ^{1}H NMR signals of the *tert*-butyl groups of (4-*tert*-butyl)tyrosine and (4-*tert*-butyl)phenylalanine are even more difficult to interrogate, as they appear near 1.3 ppm and thus invariably overlap with ^{1}H NMR signals of the protein.[23,24] Similarly, the lesser flexibility of the CF_3 groups of the genetically encodable amino acids *p*-CF_3-tyrosine and *p*-CF_3-phenylalanine results in broader ^{19}F NMR signals than observed for TFAK.[7]

An important consideration in the choice of genetically encoded amino acid is the ease and yield with which it can be installed in proteins. All three of the non-canonical amino acids used in the present work were readily installed with final protein yields of about 20 to 30 mg per liter cell culture. While the incorporation of TMSK proceeded without significant competition by canonical amino acids, the mass spectrometric analysis of the TFAK mutants indicated a degree of amber suppression by glutamine at some sites, amounting to about 10% and 20% for the mutants at sites S46 and N72, respectively. Similarly, the mass spectrum of the mutant N72TMSNK indicated misincorporation of glutamine at the amber stop codon in about 10% of the protein.

In principle, the incorporation of a non-canonical amino acid into a protein constitutes a chemical modification of the target protein, which may alter its properties. To minimize structural perturbations of the protein and maintain maximal mobility of the NMR probes, the amino acids are best installed at highly solvent-exposed sites. It is important to note that isotope labeling does not necessarily preserve the protein properties either. In particular, it is well-known that perdeuteration, which is commonly necessary for NMR studies of high molecular weight systems such as M^{pro}, significantly destabilizes the native folded state.[25–28] For large proteins, the perturbation introduced by a single non-canonical amino acid may be smaller than the cumulative effect of perdeuteration.

Comparison with Enzymatic Activity Assays. Site-directed mutagenesis has long been used to modify enzyme properties and activities, and the truncation of individual amino acid side chains by mutation to alanine is commonly used to explore the biological importance of specific sites in a target protein. Notably, the enzymatic activity of M^{pro} was maintained by all TMSK mutants. This was particularly unexpected for the N142TMSK mutant, as N142 is located centrally in the binding site of inhibitor **1**, and mutation of this residue could have completely abolished ligand binding. This example illustrates how single-site mutations may impact enzymatic activities and inhibition to a lesser degree than expected based on the space requirements of the new amino acid sidechain introduced. The alanine mutants of M^{pro} explored in the present work also maintained their enzymatic activity, which was attenuated by the presence of inhibitor **1** as expected. Activity assays of various mutants with inhibitors are thus unreliable in pinpointing the active site, although an inhibitor can be considered to bind to the active site, if the assays demonstrate competitive inhibition. The NMR probes of the present work are more broadly applicable for probing the ligand binding site, including binding sites far from the active site.

Pelitinib Activating, not Inhibiting M^{pro}. An EC_{50} value of 1.25 mM was reported for pelitinib, and the crystal structure of its complex with M^{pro} shows a binding site that does not overlap with the substrate binding site,[12] suggesting that pelitinib acts as an allosteric inhibitor. According to the crystal structure, I213 and V297 are in direct contact with pelitinib, but neither TMSK nor TFAK at these sites showed a significant response to the presence of pelitinib in our NMR experiments, suggesting that pelitinib binds elsewhere. Notably, the crystal structure of the complex displays two pelitinib molecules stacked via their quinoline rings, and each straddles not only between two M^{pro} monomers, but also

contacts a third M^{pro} molecule, suggesting that the binding site is strongly determined by the crystal lattice.

Our in vitro assays detected no inhibition by pelitinib. To the contrary, pelitinib was found to activate M^{pro} by promoting its dimerization. This suggests that the target of pelitinib in the cell based EC_{50} measurements is not M^{pro} but a yet unknown mechanism that could be of interest for antiviral drug development.

CONCLUSION

Methods that identify the binding site of a ligand easily without resorting to a full 3D structure determination continue to be a challenge, especially if the binding sites are not identical with the active site. Spectral changes observed for isotope-labelled proteins by NMR spectroscopy currently are considered among the best evidence of site-specific binding of a ligand. Such experiments require much higher concentrations of protein, usually in isotope-labelled form, much greater efforts to record and analyze the NMR spectra, and good stability of the protein to allow data recording for at least hours if not days. In this situation, the TMSK, TMSNK, and TFAK probes evaluated in the present work offer an attractive new way for sensitive site-specific detection of ligand binding. In the case of M^{pro}, binding of inhibitor **1** generated significant changes of backbone amide chemical shifts far from its binding site, indicating widespread allosteric effects.[4] The TMS and CF_3 signals of the non-canonical amino acids give a much clearer distinction between ligand binding sites near to or far from the inhibitor binding site. The distinction was also clearer than what could be obtained by enzymatic assays of alanine mutants. We note that classical approaches that combine a functional assay with different strategies of site-directed mutagenesis, such as alanine-scanning mutagenesis,[1] can be inconclusive, as the mutation of a residue in contact with the ligand may not impact ligand binding, a remote mutation site can have an impact via allosteric effects, or the protein fold may be affected by the mutation. Using NMR spectroscopy, a straightforward 1D ^1H NMR spectrum can ascertain the overall structural conservation of a protein even if individual resonances cannot be resolved.

NMR spectroscopy is widely considered a gold standard for assessing the specificity of ligand binding in solution. Most often, NMR is used to gain site-specific information, and the experiments involve labelling of the target protein with stable isotopes to improve

spectral resolution with minimal changes of protein properties, NMR resonance assignments, and identification of ligand binding sites by large chemical shift perturbations.[29] The non-canonical amino acids of the present work eliminate the resonance assignment problem. Furthermore, their signals were detectable in all mutants targeted. They open an attractive route for NMR assessments of large proteins of limited stability, especially if structural models are available that invite probing specific sites for ligand binding.

EXPERIMENTAL SECTION

Synthesis of N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine (TMSNK). *General Information.* Solvents for chromatography were technical grade. Column chromatography was performed using silica gel cartridges (Biotage, Sweden). Purity was determined by ^{1}H and ^{13}C NMR analysis. All presented compounds are ≥95% pure.

^{1}H NMR and ^{13}C NMR were recorded on a Bruker Avance 400 (400 MHz) spectrometer, using $CDCl_3$ as solvent. Data are reported in the following order: chemical shift (δ) values are reported in ppm with the solvent resonance as internal standard ($CDCl_3$: δ = 7.26 ppm for ^{1}H, δ = 77.2 ppm for ^{13}C; MeOD-d_4: δ = 3.31 ppm for ^{1}H, δ = 49.1 ppm for ^{13}C). Multiplicities are indicated as br s (broadened singlet), s (singlet), d (doublet), t (triplet), q (quartet), and m (multiplet); coupling constants (J) are given in hertz (Hz).

High-resolution mass spectra were recorded on a Waters Synapt G2 HDMS qTOF mass spectrometer connected to a Waters Acquity UPLC I Class plus LC unit. An isocratic elution mode was used to deliver the sample into the spectrometer without chromatographic separation, using methanol as the eluent at a flow rate of 0.2 mL/min.

Optical rotations were measured using an Autopol I automatic polarimeter (Rudolph Research Analytical, USA).

*N-((trimethylsilyl)methyl)-1H-imidazole-1-carboxamide (**2**).* **2** was synthesized using a protocol adapted from the literature.[6] (Trimethylsilyl)methanamine hydrochloride (**1**) (3.82 g, 27.3 mmol, purchased from Enamine, Ukraine) was suspended in CH_2Cl_2 (150 mL) under a nitrogen atmosphere, and Et_3N (7.7 mL, 55.2 mmol, 2 eq.) was added, which solubilized **1**. 1,1'-carbonyldiimidazole (5.78 g, 35.6 mmol, 1.3 eq.) was added and stirred at room temperature for 16 h. To the mixture was added H_2O (200 mL), and the solution was transferred to a separating funnel and extracted with CH_2Cl_2 (2x 100 mL), dried with

anhydrous Na_2SO_4, and concentrated to dryness by rotary evaporation. The material was purified by column chromatography (silica, 5–10% MeOH in CH_2Cl_2) to afford the title compound **2** as a white powder (4.48 g, 83%, R_f = 0.79 in 10% MeOH in CH_2Cl_2). ^{1}H NMR ($CDCl_3$, 400 MHz, 298K) δ/ppm: 8.12 (s, 1H), 7.40 (s, 1H), 7.03-6.93 (m, 2H), 2.88 (d, J = 5.7 Hz, 2H), 0.10 (s, 9H). ^{13}C NMR ($CDCl_3$, 100 MHz, 298K) δ/ppm: 149.49, 135.89, 129.88, 116.49, 31.54, -2.38.

*N^2-(tert-butoxycarbonyl)-N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine (**3**)*. **3** was synthesized by dissolving **2** (4.71 g, 23.9 mmol) in DMF (80 mL) together with Boc-Lys-OH (5.88 g, 23.9 mmol, 1 eq.) and heating to 80–90 °C for 3 h under a nitrogen atmosphere, during which the mixture became soluble. The majority of the DMF was removed by distillation under reduced pressure (or by azeotropic distillation with toluene, 5x 100 mL on the rotary evaporator). The remaining material was dissolved in CH_2Cl_2 (cloudy solution), transferred to a separating funnel, and washed with 5% aqueous LiCl (6x 100 mL, to remove residual DMF), dried with anhydrous Na_2SO_4, concentrated to dryness by rotary evaporation, and further dried under high vacuum to afford the title compound **3** as a white power (3.1 g, 34%) with no further purification required. ^{1}H NMR ($CDCl_3$, 400 MHz, 298K) δ/ppm: 5.32 (m, 1H), 4.27 (m, 1H), 3.16 (m, 2H), 2.52 (s, 2H), 1.91-1.66 (m, 2H), 1.62-1.32 (m, 13H), 0.09 (s, 9H). ^{13}C NMR ($CDCl_3$, 100 MHz, 298K) δ/ppm: 175.59, 160.81, 155.82, 79.90, 53.47, 40.44, 32.42, 30.83, 29.61, 28.52, 22.35, -2.64.

*TMSNK hydrochloride (**4**)*. **4** was synthesized by suspending **3** (3.1 g, 8.3 mmol) in Et_2O (100 mL) and cooling on ice to 0 °C under a nitrogen atmosphere. 4 M HCl in 1,4-dioxane (13.5 mL, 54 mmol, 6.5 eq.) was added dropwise over 5 minutes and allowed to warm to room temperature with stirring for 3.5–4.5 h, after which the solution was concentrated to dryness. The material was triturated with 1:1 Et_2O: *n*-hexane and stirred vigorously for 1 h, with the resulting white powder filtered. This was repeated for three cycles to afford a free flowing (not sticky) white powder which was further dried under high vacuum and characterized to be the title compound **4** (2.3 g, 89 %). ^{1}H NMR (MeOD-d_4, 400 MHz, 298K) δ/ppm: 3.97 (t, J = 6.3 Hz, 1H), 3.21 (m, 2H), 2.65 (s, 2H), 2.06-1.83 (m, 2H), 1.68-1.39 (m, 4H), 0.10 (s, 9H). ^{13}C NMR (MeOD-d_4, 100 MHz, 298K) δ/ppm: 171.72, 161.47, 53.76, 41.42, 31.91, 31.14, 30.33, 23.23, -2.76. LR-ESI-MS (positive mode) for $[C_{11}H_{26}N_3O_3Si]^+$ $[M+H]^+$: Calc: 276.2. Found: 276.2. HR-ESI-TOF-MS (positive mode)

for $[C_{11}H_{25}N_3O_3SiNa]^+$ $[M+Na]^+$: Calc: 298.1563. Found: 298.1567. $[a]_D^{25}$: +11.3, *c* 1.0, MeOH.

Screening of functional G1PylRS enzymes recognizing TMSNK. Aminoacyl-tRNA synthetases (RS) specific for TMSNK were selected from a library of RS mutants based on the pyrrolysyl-tRNA synthetase (PylRS) from the methanogenic archaeon ISO4-G1, using fluorescence-activated cell sorting (FACS) as described previously.[6,21,22] To carry out the selection, the library plasmid pBK-G1RS was transformed into *E. coli* DH10B cells harboring the selection plasmid pBAD-H6RFP.[30] Following recovery from transformation, the culture was directly inoculated into a flask with 25 mL LB medium containing 100 mg/L carbenicillin and 50 mg/L kanamycin and supplied with 0.4% L-arabinose and 2 mM TMSNK, which served as the sample for the first round of positive selection (1P+). Overnight expression at 37 °C led to a readily detectable level of red fluorescence protein (RFP) expression. Cells were resuspended in 7.5 mL PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) after harvesting. A 100-fold dilution yielded a concentration suitable for cell sorting by FACS on an Aria Fusion high-speed cell sorter (BD Biosciences, USA; Figure S9).

Cells with high RFP levels were selected from the 1P+ sample (1.2% of the total population, as indicated by rectangles in Figure S9), collecting 6.0×10^5 cells in 60 minutes. The cells collected were subjected to a following round of negative selection without the addition of TMSNK and regrown as sample 2N–, from where cells with low RFP expression levels (31.1%) were collected (8×10^5 cells in 15 minutes). These cells were aliquoted to inoculate media with positive (3P+) and negative (3P–) conditions. The RFP-positive cells (9.0%) from the 3P+ round were collected (4.0×10^5 cells in 25 minutes) and recovered under negative condition to obtain 4N–. Following sorting (6.0×10^5 cells in 10 minutes), cells showing the lowest level of RFP fluorescence (41.1%) were selected from the 4N– sample. They were aliquoted to be recovered as 5P+ and 5P–, respectively. The cell population with high RFP fluorescence in the 5P round was 33.9% with TMSNK provided compared to 0.3% without TMSNK, indicating the successful accumulation of active G1PylRS variants specific for TMSNK. 2.0×10^5 cells collected in 35 minutes from the top 2.2% RFP fluorescent cells of the 5P+ sample were recovered for storage. An aliquot of 2,000 cells was allowed to recover on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin, and individual clones were analyzed using 96-well

plates. 60 candidates were inoculated into both positive (with 2 mM TMSNK) and negative (without TMSNK) growth conditions. The fluorescence level was measured after expression overnight using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland) and normalized by the OD_{600} of the cell culture. Seven candidates with the highest RFP level in the positive condition were chosen for sequencing. Three individually different sequences were found (Table S1), with one mutation set (V167G, Y204W) identified three times, one mutation set (V167G, Y204F, A221C) identified two times, and another mutation set (V167G, Y204W, A221C) identified two times. The clone producing the highest fluorescence (TMSNK11) was found to carry the mutations V167G and Y204W. This G1PylRS mutant was used in the subsequent experiments and referred to as G1TMSNKRS. The plasmids pRSF-G1TMSNKRS, pRSF-ChPylTMSK, and pRSF-G1TFAKRS for the incorporation of TMSNK, TMSK, and TFAK, respectively, are available from Addgene (Teddington, UK) (plasmids # 198323, 163915, and 177311).

Protein production. *Plasmid construction.* The gene corresponding to M^{pro} [14] was cloned in between the *Nde*I and *Xho*I sites of the modified T7 vector pCDF.[6] The M^{pro} construct contained the natural M^{pro} cleavage-site (SAVLQ↓SGFRK; arrow indicating the cleavage site) at the N-terminal end and a modified PreScission cleavage site (SGVTFQ↓GP) followed by a His_6 tag at the C-terminal end. Point mutagenesis introduced an amber stop codon (TAG) at selected sites for incorporation of TMSK, TMSNK, and TFAK and a GCG codon for incorporation of Ala using overlapping primers. A mutant T4 DNA polymerase was used in in-house RQ-SLIC and QuikChange protocols for cloning and mutagenesis, respectively.[31]

Protein expression and purification. E. coli B-95.ΔA cells[32] were co-transformed with pRSF-ChPylTMSKRS,[6] pRSF-G1TMSNKRS (Table S1), pRSF-G1TFAKRS,[7] and the respective pCDF plasmid to produce M^{pro} mutants with TMSK, TMSNK, and TFAK. The cells were grown at 37 °C in LB medium containing 50 mg/L kanamycin and 50 mg/L spectinomycin. Alanine mutants were produced in *E. coli* BL21 DE3 transformed with the requisite pCDF-M^{pro} plasmid, growing the cells at 37 °C in LB medium containing 50 mg/L spectinomycin. Overnight cultures were inoculated into fresh terrific broth (TB) medium (1:100 dilution) supplemented with 50 mg/L kanamycin and 50 mg/L spectinomycin (TMSK and TMSNK mutants) and fresh Luria-Bertani (LB) medium (1:100 dilution) supplemented with 50 mg/L spectinomycin (alanine mutants) or 50 mg/L kanamycin and

50 mg/L spectinomycin (TFAK). The cells were grown at 37 °C to an OD_{600} of 0.6–1. Expression of mutants was initiated by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) and TMSK (1 mM; BAPEKS, Riga, Latvia), TMSNK (2 mM), or TFAK (20 mM, AK Scientific, USA). Expression of the alanine mutants was induced by 1 mM IPTG. Protein expression was conducted by incubation at room temperature overnight.

Cells were harvested by centrifugation at 5,000 *g* for 15 minutes. Following resuspension in buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl), the cells were lysed using an Avestin Emulsiflex C5 (Avestin, Canada) (two passes using 10,000–15,000 psi). The cell lysates were centrifuged for 1 h at 30,000 *g*. The supernatant was loaded onto a 1 mL His GraviTrap TALON® column (Cytiva, USA) equilibrated with buffer A. The column was washed with 20 column volumes buffer B (same as buffer A but with 5 mM imidazole), and the protein was eluted with 5 column volumes buffer C (same as buffer A but with 300 mM imidazole). The fractions were analyzed by 12 % SDS-PAGE. Following the purification, the buffer was exchanged to the NMR buffer (20 mM HEPES-KOH, pH 7.0, 150 mM NaCl, 1 mM DTT, 1 mM EDTA using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with a molecular weight cut-off of 10 kDa. All experiments of the present work used samples that retained the C-terminal His_6 tag, and their correctness was assessed by mass spectrometry using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific, USA) coupled with an UltiMate S4 3000 UHPLC (Thermo Scientific, USA).

Protein concentrations were determined by measuring the absorbance at 280 nm, using e = 33,640 M^{-1} cm^{-1}. Yields of purified protein per liter cell culture were 16 – 24 mg/L for the TMSK mutants, 26 – 33 mg/L for the TFAK mutants, and 17 – 19 mg/L for the TMSNK mutants except for the mutants at sites 191 and 186, which yielded only about 1 mg/L.

NMR experiments. 1D ^{1}H NMR and 1D ^{19}F NMR spectra were recorded at 25 °C using 3 mm NMR tubes on a 600 MHz Bruker Avance NMR spectrometer equipped with a 5 mm TCI cryoprobe that allows tuning of the ^{1}H coil to ^{19}F. The ^{1}H and ^{19}F NMR experiments used 30 mM and 100 mM solutions of M^{pro}, respectively. 10% D_2O was added to provide a lock signal. The ^{1}H NMR spectra were recorded using the jump-return sequence,[33] using 256 scans per spectrum. Each ^{19}F NMR spectrum was recorded using 256 scans, an acquisition time of 0.29 s, and a recording time of 5 minutes. The ^{19}F NMR

spectra were calibrated relative to internal trifluoroethanol at -76.45 ppm.[8] The titration experiments were conducted with 0.2 mL protein solutions and stock solutions of ligands in DMSO that, for the 1:1 titration ratios, diluted the protein by 3% in the case of the TMSK and TMSNK mutants and 4% in the case of the TFAK mutants.

Activity assay. The activity assays were based on published M^{pro} Förster resonance energy transfer (FRET) assays[14,34] and conducted as described previously.[35] The assay was conducted with a total volume of 100 µL in 96-well plates (black, polypropylene, U-bottom; Greiner Bio-One, Austria) using 20 mM Tris-HCl pH 7.3 buffer containing 100 mM NaCl, 1 mM EDTA, and 1 mM DTT. Samples of recombinant wild-type and mutant M^{pro} were incubated at 25 nM concentration with ligand compound for 10 minutes at 37 °C before the addition of the fluorogenic substrate (DABCYL)-KTSAVLQ↓SGFRKM-E(EDANS)-NH_2 (25 µM; Mimotopes, Australia) started the enzymatic reaction. A fluorophotometer (Infinite 200 PRO M Plex; Tecan, Switzerland) was used to monitor the initiated enzymatic reaction ($\lambda_{excitation}$ = 340 nm, $\lambda_{emission}$ = 490 nm) for 5 minutes. All experiments were performed in triplicate. The initial velocities of control reactions with M^{pro} variant and FRET substrate in the absence of inhibitor were defined as 100% enzymatic activity, respectively. The experimental data sets were analyzed and visualized with Prism 9.2 (GraphPad Software, USA). The cyclic peptide inhibitor **1** was as reported previously.[4] Calpeptin and pelitinib were purchased from MedChemExpress (USA).

Gel filtration experiments. To a solution of M^{pro}-His_6 (550 µL, 1 µM) in buffer (20 mM Tris-HCl, 150 mM NaCl, pH 8.0) was added a solution of calpeptin or pelinitib (2.5 mM in DMSO) to achieve a final inhibitor concentration of 2 µM or 10 µM as specified. The solution was incubated at room temperature for 5 minutes before the samples were subjected to SEC (Superdex® 75 10/300 Increase GL) on an ÄKTA system fitted with MALS, UV absorbance, and refractive index (dRI) detectors, eluting with running buffer (20 mM Tris.HCl, 150 mM NaCl, pH 8.0) at 0.5 mL/min. MALS, UV, and dRI data were collected and analyzed using the ASTRA software (Wyatt) with molecular weight determinations carried out according to the Debye-Zimm model. Uncertainty in the molecular weight determination from this system is estimated to be ±10%.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website. NMR spectra of TMSNK and synthetic intermediates (Figures S1–S6); mass spectra of TMSNK (Figures S7–S8); selection of aminoacyl-tRNA synthetase mutants from a G1 PylRS library (Figure S9); mutations found in selected G1 PylRS variants (Table S1); intact protein mass spectrometric analysis of TMSK, TFAK, and TMSNK mutants (Figures S10–S12); ^{19}F NMR spectra of TFAK mutants titrated with DMSO (Figure S13); ^{19}F NMR spectra of TFAK mutants titrated with inhibitor **1** purified with acetonitrile/formic acid (Figure S14); stability of TMSNK versus TMSK mutants monitored by ^{19}F NMR (Figure S15).

AUTHOR INFORMATION

Corresponding Author

*** Gottfried Otting** – Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia; Email: gottfried.otting@anu.edu.au

Authors

Kasuni B. Ekanayake – Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Mithun C. Mahawaththa – Australian Research Council Centre of Excellence for Innovations in Peptide and 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

Elwy H. Abdelkader – Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Josemon George – Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Rhys B. Murphy – Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Sven Ullrich – Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Sarah E. Fry – Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science and School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia

Jason Johansen-Leete – Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science and School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia

Richard J. Payne – Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science and School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia

Christoph Nitsche – Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Thomas Huber – Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Author Contributions

K.B.E. produced and mutated the proteins, measured the NMR spectra, analyzed the data, and wrote the first draft of the manuscript. M.C.M. designed the mutants. H.Q., E.H.A., and T.H. developed the genetic encoding system of TMSNK. J.G. and R.B.M. synthesized TMSNK. S.U. performed the enzyme activity tests. S.E.F. and R.J.P. provided the cyclic peptide **1**. J.J.-L. performed the final SEC-MALS experiments. C.N. supervised the synthesis of TMSNK and the activity assays. G.O. designed the research and wrote the final manuscript.

Funding

Research funded by the Australian Research Council for a Laureate Fellowship to G.O. (FL170100019), a Future Fellowship to C.N. (FT220100010), projects DP200100348 and DP21010088, and through the Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012 to R.J.P. and G.O.).

ACKNOWLEDGMENTS

We thank Mr. Richard Morewood for the synthesis of an initial batch of TFAK and Dr. Harpreet Vohra and Michael Devoy at the John Curtin School of Medical Research, The Australian National University, for technical support on FACS experiments.

ABBREVIATIONS USED

CoV, corona virus; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; HPLC, high performance LC; IPTG, isopropyl-β-D-thiogalactopyranoside; LB, Luria-Bertani; LC, liquid chromatography; M^{pro}, SARS-CoV-2 main protease; MALS, multi-angle light scattering; NMR, nuclear magnetic resonance; SARS, severe acute respiratory syndrome; TFAK, N^6-trifluoroacetyl-L-lysine; TMSK, N^6-(((trimethylsilyl)-methoxy)carbonyl)-L-lysine; TMSNK, N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine.

REFERENCES

(1) Cunningham, B. C.; Wells, J. A. High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science* **1989**, *244*, 1081–1085.

(2) Kim, J.; Wess, J.; Rhee, A. M. V.; Schöneberg, T.; Jacobson, K. A. Site-Directed Mutagenesis Identifies Residues Involved in Ligand Recognition in the Human A_{2a} Adenosine Receptor. *J. Biol. Chem*. **1995**, *270*, 13987–13997.

(3) Chen, G.; Dubrawsky, I.; Mendez, P.; Georgiou, G.; Iverson, B. L. *In vitro* Scanning Saturation Mutagenesis of all the Specificity Determining Residues in an Antibody Binding Site. *Protein Eng. Des. Sel*. **1999**, *12*, 349–356.

(4) Johansen-Leete, J.; Ullrich, S.; Fry, S. E.; Frkic, R.; Bedding M. J.; Aggarwal, A.; Ashhurst, A. S.; Ekanayake, K. B.; Mahawaththa, M. C.; Sasi, V. M.; Luedtke, S.; Ford, D. J.; O'Donoghue, A. J.; Passioura, T.; Larance, M.; Otting, G.; Turville, S.; Jackson, C. J.; Nitsche, C.; Payne, R. J. Antiviral Cyclic Peptides Targeting the Main Protease of SARS-CoV-2. *Chem. Sci*. **2022,** *13*, 3826–3836.

(5) Zhang, F.; Zhou, Q.; Yang, G.; An, L.; Li, F.; Wang, J. A. Genetically Encoded ^{19}F NMR Probe for Lysine Acetylation. *Chem. Commun*. **2018**, *54*, 3879−3882.

(6) Abdelkader, E. H.; Qianzhu, H.; Tan, Y. J.; Adams, L. A.; Huber, T.; Otting, G. Genetic Encoding of N^6-(((Trimethylsilyl)methoxy)carbonyl)-L-lysine for NMR Studies of

Protein−Protein and Protein−Ligand Interactions. *J. Am. Chem. Soc*. **2021**, *143*, 1133−1143.

(7) Orton, H. W.; Qianzhu, H.; Abdelkader, E. H.; Habel, E. I.; Tan, Y. J.; Frkic, R. L.; Jackson, C.; Huber, T.; Otting, G. Through-Space Scalar $^{19}F–^{19}F$ Couplings between Fluorinated Noncanonical Amino Acids for the Detection of Specific Contacts in Proteins. *J. Am. Chem. Soc.* **2021**, *143*, 19587–19598.

(8) Rosenau, C. P.; Jelier, B. J.; Gossert, A. D.; Togni A. Exposing the Origins of Irreproducibility in Fluorine NMR Spectroscopy. *Angew. Chem. Int. Ed*. **2018**, *57*, 9528–9533.

(9) Didenko, T.; Liu, J. J.; Horst, R.; Stevens, R. C.; Wüthrich, K. Fluorine-19 NMR of Integral Membrane Proteins Illustrated with Studies of GPCRs. *Curr. Opin. Struct. Biol*. **2013**, *23*, 740–747.

(10) Picard, L.-P.; Prosser R. S. Advances in the Study of GPCRs by ^{19}F NMR. *Curr. Opin. Struct. Biol*. **2021**, *69*, 169–176.

(11) Yang, H.; Yang, J. A Review of the Latest Research on M^{pro} Targeting SARS-COV Inhibitors. *RSC Med. Chem*. **2021**, *12*, 1026–1036.

(12) Günther, S.; Reinke P. Y. A.; Fernández-García Y.; Lieske, J.; Lane,T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J. -D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Perez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H., Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlünzen, F.; Xavie, P. L.; Werner, N.; Andaleeb, H.; Ullah, N.;Falke, S.; Srinivasan, V.; França, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitseva-Doyle, J. J.; Knoska, J.; Peña-Murillo, G. E.; Mashhour, A. R.; Hennicke, V.; Fischer, P.; Hakanpää, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; Stetten D. V.; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia-Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E. -C.; Mehrabi, P.; Karnicar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.;

Chapman, H. N.; Pearson, A. R.; Betze, C.; Meents A. X-ray Screening Identifies Active Site and Allosteric Inhibitors of SARS-CoV-2 Main Protease. *Science* **2021**, *372*, 642–646.

(13) Silvestrini, L.; Belhaj, N.; Comez, L.; Gerelli, Y.; Lauria, A.; Libera, V.; Mariani, P.; Marzullo, P.; Ortore, M. G.; Piccionello A. P.; Petrillo, C.; Savini, L.; Paciaroni, A.; Spinozzi, F. The Dimer-Monomer Equilibrium of SARS-CoV-2 Main Protease is Affected by Small Molecule Inhibitors. *Sci. Rep.* **2021**, *11*, 9283.

(14) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020**, *368*, 409–412.

(15) El-Baba, T. J.; Lutomski, C. A.; Kantsadi, A. L.; Malla, T. R.; John, T.; Mikhailov, V.; Bolla, J. R.; Schofield, C. J.; Zitzmann, N.; Vakonakis, I.; Robinson, C. V. Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angew. Chem. Int. Ed.* **2020**, *59*, 23544–23548.

(16) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of M^{pro} From SARS-CoV-2 and Discovery of its Inhibitors. *Nature* **2020**, *582*, 289–293.

(17) Sasi, V. M.; Ullrich, S.; Ton, J.; Fry, S. E.; Johansen-Leete, J.; Payne, R. J.; Nitsche, C.; Jackson, C. J. Predicting Antiviral Resistance Mutations in SARS-CoV2 Main Protease with Computational and Experimental Screening. *Biochemistry* **2022**, *61*, 2495–2505.

(18) 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.

(19) Ge, H.; Wang, H.; Pan, B.; Feng, D.; Guo, C.; Yang, L.; Liu, D.; Wüthrich, K. G Protein-coupled Receptor (GPCR) Reconstitution and Labeling for Solution Nuclear Magnetic Resonance (NMR) Studies of the Structural Basis of Transmembrane Signaling. *Molecules* **2022**, *27*, 2658.

(20) Wang, X.; Liu, D.; Shen, L.; Li, F.; Li, Y.; Yang, L.; Xu, T.; Tao, H.; Yao, D.; Wu, L.; Hirata, K.; Bohn, L. M.; Makriyannis, A.; Liu, X.; Hua, T.; Liu, Z.-J.; Wang, J. A Genetically Encoded F-19 NMR Probe Reveals the Allosteric Modulation Mechanism of Cannabinoid Receptor 1. *J. Am. Chem. Soc*. **2021**, *143*, 16320–16325.

(21) Qianzhu, H.; Welegedara, A. P.; Williamson, H.; McGrath, A. E.; Mahawaththa, M. C.; Dixon, N. E.; Otting, G.; Huber, T. Genetic Encoding of *para*-Pentafluorsulfanyl Phenylalanine: A Highly Hydrophobic and Strongly Electronegative Group for Stable Protein Interactions. *J. Am. Chem. Soc.* **2020**, *142*, 17277–17281.

(22) 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*, 44–49.

(23) 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.

(24) Chen, W. N.; Kuppan, K. V.; Lee, M.; 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.

(25) Hattori, A.; Crespi, H. L.; Katz, J. J. Effect of Side-Chain Deuteration on Protein Stability. *Biochemistry* **1965**, *4*, 1213−1225.

(26) Brockwell, D.; Yu, L.; Cooper, S.; Mccleland, S.; Cooper, A.; Attwood, D.; Gaskell, S. J.; Barber, J. Physicochemical Consequences of the Perdeuteriation of Glutathione S-Transferase from *S. japonicum*. *Prot. Sci.* **2001**, *10*, 572–580.

(27) 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.

(28) 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. *Prot. Sci*. **2020**, *29*, 1641−1654.

(29) Nitsche, C.; Otting, G. NMR Studies of Ligand Binding. *Curr. Opin. Struct. Biol.* **2018**, *48*, 16–22.

(30) Abdelkader, E. H.; Qianzhu, H.; George, J.; Frkic, R. L.; Jackson, C. J.; Nitsche, C.; Otting, G.; Huber, T. Genetic Encoding of Cyanopyridylalanine for In-Cell Protein Macrocyclization by the Nitrile-Aminothiol Click Reaction. *Angew. Chem. Int. Ed.* **2022**, *61*, e202114154.

(31) Qi, R.; Otting, G. Mutant T4 DNA Polymerase for Easy Cloning and Mutagenesis. PLOS One **2019**, *14*, e0211065.

(32) 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.

(33) Plateau, P.; Guéron, M. Exchangeable Proton NMR Without Base-Line Distortion, Using New Strong-Pulse Sequences. *J. Am. Chem. Soc*. **1982**, *104*, 7310–7311.

(34) Zhu, L.; George, S.; Schmidt, M. F.; Al-Gharabli, S. I.; Rademann, J.; Hilgenfeld, R. Peptide Aldehyde Inhibitors Challenge the Substrate Specificity of the SARS-Coronavirus Main Protease. *Antivir. Res*. **2011**, *92*, 204–212.

(35) Ullrich, S.; Ekanayake, K. B.; Otting, G.; Nitsche, C. Main Protease Mutants of SARS-CoV-2 Variants Remain Susceptible to Nirmatrelvir. *Bioorg. Med. Chem. Lett*. **2022**, 128629.