The FEBS Journal

FEBS PRESS science publishing by scientists

Comparative evaluation of noncanonical amino acids as site-specific NMR probes for the complex of *E. coli* SSB with single-stranded DNA without isotope labelling

Sreelakshmi Mekkattu Tharayil[1], Haocheng Qianzhu[1], Elwy H. Abdelkader[2], Adarshi P. Welegedara[2], Josemon George[1], Christoph Nitsche[1], Thomas Huber[1] and Gottfried Otting[2]

1 Research School of Chemistry, The Australian National University, Canberra, Australia

2 Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, The Australian National University, Canberra, Australia

Keywords

^{19}F-NMR spectroscopy; *E. coli* single-stranded DNA-binding protein (SSB); genetic encoding; noncanonical amino acids; nuclear magnetic resonance (NMR)

Correspondence

G. Otting, Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, The Australian National University, Canberra, ACT 2601, Australia

Tel.: +61 2 612 56507

E-mail: gottfried.otting@anu.edu.au

Sreelakshmi Mekkattu Tharayil and Haocheng Qianzhu contributed to the work equally.

(Received 15 August 2025, revised 19 November 2025, accepted 30 January 2026)

doi:10.1111/febs.70446

The binding of tetrameric *E. coli* single-stranded DNA-binding protein (SSB) to single-stranded DNA (ssDNA) was investigated using genetically encoded noncanonical amino acids (ncAA) as site-specific probes for detection by nuclear magnetic resonance (NMR) spectroscopy. Under the conditions used (300 mM NaCl, pH 7.2), the NMR spectra confirmed the equivalence of the monomeric subunits in the absence of ssDNA. Most of the probes responded to the binding of ssDNA by changes in chemical shifts and line width and distinguished between the presence of segments of cytidine versus thymidine. Although ssDNA-binding breaks the fourfold symmetry of the SSB tetramer, the probes sensed closely similar chemical environments in all four monomeric subunits. By comparing the performance of twelve different NMR-active ncAAs, this work identified N^6-trifluoroacetyl-L-lysine (TFAK) as the ncAA sensing different ssDNAs with the best spectral resolution. In addition, we report aminoacyl-tRNA synthetases for the genetic encoding of 3,5-difluoro-L-tyrosine (3,5-diFTyr), 2,6-difluoro-L-tyrosine (2,6-diFTyr), and *m*CF_3-phenylalanine. The SSB construct was sensitive to precipitation under NMR conditions. The fluorinated ncAAs altered the rates of precipitation which varied even between fluorotryptophan isomers installed at a barely solvent-accessible site. Nonetheless, the NMR-active ncAAs proved suitable for probing a marginally stable protein system of ca. 100 kDa molecular weight without isotope labelling and at low concentration. The current data suggest that ^{19}F spins attached to flexible solvent-exposed amino acid side chains guard better against protein precipitation than fluorinated aromatic amino acids despite the latter being more attractive for their close structural similarity to their canonical amino acid counterparts.

Abbreviations

2,6-diFPhe, 2,6-difluoro-L-phenylalanine; 2,6-diFTyr, 2,6-difluoro-L-tyrosine; 3,5-diFPhe, 3,5-difluoro-L-phenylalanine; 3,5-diFTyr, 3,5-difluoro-L-tyrosine; 4FTrp, 4-fluoro-L-tryptophan; 5FTrp, 5-fluoro-L-tryptophan; 6FTrp, 6-fluoro-L-tryptophan; 7FTrp, 7-fluoro-L-tryptophan; FACS, fluorescence-activated cell sorting; *m*CF_3Phe, 3-(trifluoromethyl)-L-phenylalanine; ncAA, noncanonical amino acid; NMR, nuclear magnetic resonance; SSB, single-stranded DNA-binding protein; ssDNA, single-stranded DNA; TFAK, N^6-trifluoroacetyl-L-lysine; TMSK, N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine; TMSNK, N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine.

Introduction

Genetic encoding of noncanonical amino acids (ncAA) has emerged as a convenient technique for probing large protein systems site-specifically by NMR spectroscopy using straightforward one-dimensional NMR experiments. The approach affords maximal spectral resolution and selectivity, delivering resonance assignments without the need for time-consuming experiments such as multidimensional NMR or site-directed mutagenesis. The most cost-effective strategies use ncAAs which function without chemical protection groups and avoid expensive isotope labeling by relying on either a very intense ^{1}H NMR signal [1–4] or fluorine-containing ncAAs for detection by ^{19}F NMR [5–11]. The ncAAs are typically encoded by the amber stop codon and installed during protein expression.

To date, most ncAAs designed for NMR detection have been demonstrated with different protein systems, making it difficult to compare the performance of the ncAAs. Exceptions include flaviviral proteases (about 27 kDa molecular weight) probed with sets of differently fluorinated tryptophan and phenylalanine analogues [9–11]. Among these, the Zika virus protease has also been explored with genetically encoded *tert*-butyltyrosine and (4-trimethylsilyl)phenylalanine, but the ^{1}H-NMR signals of these amino acids were only incompletely resolved from the background signals of the protein [2]. In addition, genetically encoded N^{6}-(((trimethylsilyl) methoxy)carbonyl)-L-lysine (TMSK), N^{6}-(((trimethylsilyl)methyl)carbamoyl)-L-lysine (TMSNK), and N^{6}-trifluoroacetyl-L-lysine (TFAK) were compared in the SARS-CoV-2 main protease (M^{pro}), demonstrating facile detection of their respective ^{1}H- and ^{19}F-NMR signals without isotope labelling in this 67 kDa homodimer [12].

In the present work, we used the ~ 100 kDa complex of the tetrameric *Escherichia coli* single-stranded DNA-binding protein (SSB) [13] with single-stranded DNA (ssDNA) to compare the performance of twelve different ncAAs in detecting the binding of ssDNA with different nucleotide sequences by NMR spectroscopy. Investigating isotope-labeled SSB using conventional multidimensional NMR techniques is difficult and costly, as the C-terminal domain of SSB is prone to digestion at neutral pH, especially in the absence of salt [14]. In previous work, we investigated the SSB−ssDNA complex by installing custom-synthesized ^{13}C-labeled and deuterated *tert*-butyltyrosine for background-free NMR detection [15]. Here, we exclusively used protein, ncAAs and ssDNA at natural isotopic abundance in the search for cost-effective alternatives.

SSB forms a ~ 75 kDa homotetramer. Each monomer comprises an N-terminal domain that adopts a conserved oligonucleotide/oligosaccharide-binding (OB) fold responsible for high-affinity binding to ssDNA [16] and a C-terminal tail that is intrinsically disordered [17,18]. ssDNA binds to SSB by wrapping around the tetramer [19,20]. Depending on salt concentration, ssDNA such as poly(dT) can bind in two distinct binding modes referred to as $(SSB)_{35}$ and $(SSB)_{65}$, where the subscripts denote the number of nucleotides occluded per SSB tetramer [21–23]. In the $(SSB)_{65}$ binding mode, all four subunits of the tetramer are in contact with ssDNA. The present work was conducted to investigate the $(SSB)_{65}$ binding mode.

A single X-ray structure is available of *E. coli* SSB in complex with ssDNA. In this structure, the ssDNA consists of two molecules of 35-mer oligomers made of cytidines, dC_{35}, bound to the SSB tetramer (PDB ID: 1EYG [19]; Fig. 1A). The complex reflects the $(SSB)_{65}$ mode. The dC_{35} molecules contact the monomeric subunits of the tetramer in similar but not identical ways, with electron density missing for some of the nucleotides. The crystal contained only the ssDNA-binding OB domain of the protein (residues 1–115) as the disordered C-domain (residues 116–177) had been removed by digestion with chymotrypsin. To rationalize the extraordinary cooperativity and flexibility of SSB binding to ssDNA, different binding modes have been proposed, including the possibility that ssDNA traces the nucleotide-binding groove of the SSB monomers in different directions [20]. The present work probed the equivalence of the monomers in the SSB tetramer in the $(SSB)_{65}$ mode with the help of different genetically encoded NMR probes that minimally perturb the 3D structure of the tetramer.

Fig. 2 shows the ncAAs used. They comprise ncAAs designed for detection by ^{1}H- and ^{19}F-NMR. The trimethylsilyl groups of TMSK and TMSNK contain nine equivalent protons that produce a narrow and tall singlet in the ^{1}H-NMR spectrum near 0 ppm, where proteins show few other signals [2,4,12]. A second group comprises fluorinated amino acids for detection by ^{19}F-NMR spectroscopy. These probes represent fewer nuclear spins but, as canonical amino acids and nucleotides do not contain fluorine, the ^{19}F-NMR spectra are free of background signals, facilitating the detection of minor species. Furthermore, the chemical shift range of ^{19}F is much greater than that of ^{1}H spins, amplifying the size of chemical shift changes in response to ligand binding [10,11,24,25]. We report the purity and yields of SSB mutants produced with the different ncAAs, their NMR spectra, and their capacity to report the binding of ssDNA.

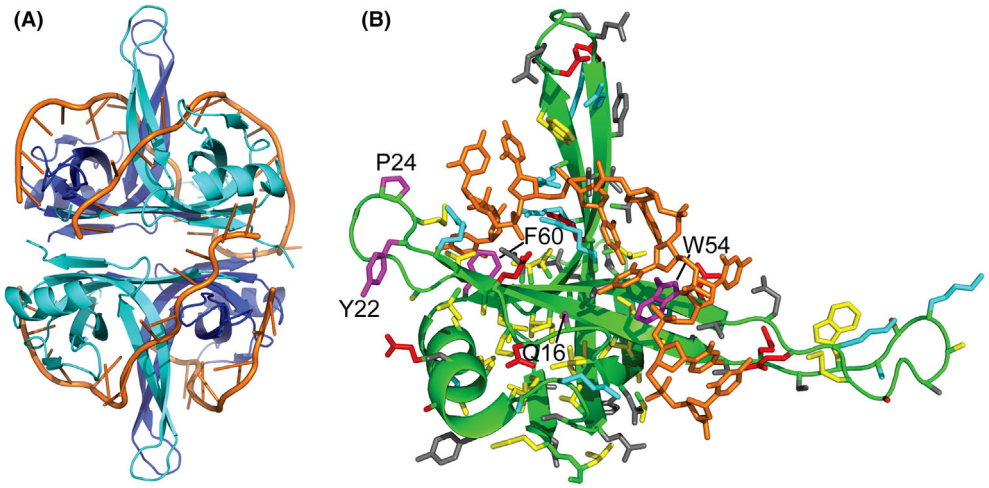

Fig. 1. Sites selected for incorporating ncAAs in the SSB tetramer (PDB ID: 1EYG [19]). (A) Cartoon representation of the SSB tetrameric structure with the dC_{35} oligomers shown in orange. (B) Structure of the monomeric subunit highlighting the positions of the ncAAs relative to the ssDNA-binding groove, showing the nucleotide and amino acid side chains in stick representation. Color code: protein backbone (green), ssDNA (orange), hydrophobic side chains (yellow), hydrophilic side chains (gray), positively charged side chains (blue), negatively charged side chains (red). The sites selected in the present work for installing ncAAs are shown in magenta and labeled with the amino acid code and sequence number in the wild-type protein. The side chain of Q16 is solvent exposed and no electron density beyond the C^{β} atom was reported in chain A. Figure prepared with PYMOL [60].

Results

Selection of mutation sites and nucleotide sequences

We inspected the crystal structure (PDB code: 1EYG) [19] to introduce ncAAs at sites, where the non-native chemical groups of the ncAA were (a) unlikely to perturb the three-dimensional quaternary structure of the protein, (b) unlikely to clash with the ssDNA, (c) sufficiently close to the ssDNA to sense its presence, and (d) sufficiently remote from the interfaces between the monomers to allow assigning the observed effects predominantly to single monomers. When the ncAA was larger than the amino acid in the wild-type protein (TMSK, TMSNK, TFAK), the sites chosen were highly solvent-exposed. Sites of lesser solvent exposure were chosen when the ncAA was structurally similar to the canonical amino acid substituted (fluorinated analogues of tyrosine, phenylalanine, and tryptophan).

For example, the side chain of Q16 is highly solvent-exposed and the crystal structure reports no electron density beyond the C^{β} atom in two of the four monomers. Yet, a TMS or TFA group at the end of a long and flexible lysine side chain may contact the bound ssDNA (Fig. 1B). We therefore selected this site for installing the aliphatic amino acids TMSK, TMSNK, and TFAK. The same reasoning applies to P24, which is in a loop region with high solvent exposure. Y22 is in the same loop as P24 but pointing away from the ssDNA. A fluorinated tyrosine version at this site was expected to sense the presence of ssDNA indirectly rather than by direct contact. In contrast, the crystal structure reports W54 and F60 to engage in direct contacts with the ssDNA, which should greatly change the chemical shifts of fluorinated analogues without generating severe steric clashes.

All NMR measurements were conducted at a high-salt concentration (300 mM NaCl) to promote the $(SSB)_{65}$ binding mode [16] and create similar chemical environments for all four monomers in the tetramer. The ssDNA segments chosen were the 66-mers dT_{66}, $dT_{1–36}C_{37–66}$, where the first 36 nucleotides are thymidines and the following 30 nucleotides are cytidines, and $dT_{1–50}C_{51–66}$.

Protein production, yields, and purity

Each of the ncAAs shown in Fig. 2 was installed in SSB by genetic encoding, using ribosomal readthrough of an amber stop codon by a suppressor-tRNA loaded with the requisite ncAA by a mutant aminoacyl-tRNA synthetase (RS). For 3,5-diFTyr and 2,6-diFTyr, new aminoacyl-tRNA synthetases were identified to achieve genetic encoding. For protein purification, the SSB construct used contained a TEV cleavage site followed by a His_6-tag at the C terminus. Following cleavage with TEV protease, this added the peptide ENLYFQ

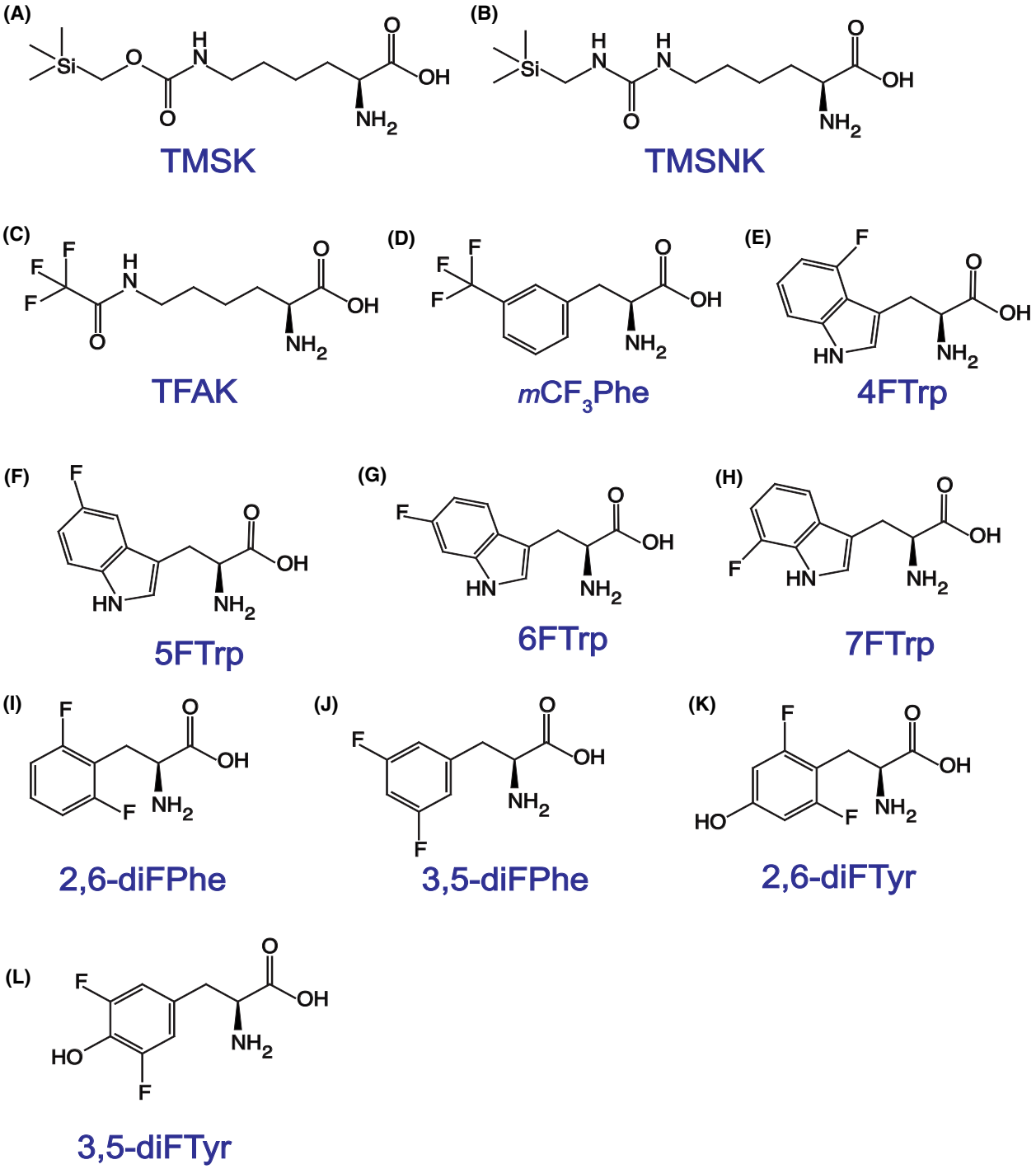

Fig. 2. Chemical structures of ncAAs used in the present work. (A) N^6-(((trimethylsilyl)methoxy)carbonyl)-L-lysine (TMSK), (B) N^6-(((trimethylsilyl)methoxy)carbamoyl)-L-lysine (TMSNK), (C) N^6-trifluoroacetyl-L-lysine (TFAK), (D) 3-(trifluoromethyl)-L-phenylalanine (mCF$_3$Phe), (E) 4-fluoro-L-tryptophan (4FTrp), (F) 5-fluoro-L-tryptophan (5FTrp), (G) 6-fluoro-L-tryptophan (6FTrp), (H) 7-fluoro-L-tryptophan (7FTrp), (I) 2,6-difluoro-L-phenylalanine (2,6-diFPhe), (J) 3,5-difluoro-L-phenylalanine (3,5-diFPhe), (K) 2,6-difluoro-L-tyrosine (2,6-diFTyr), and (L) 3,5-difluoro-L-tyrosine (3,5-diFTyr).

Table 1. Mutation sites, noncanonical amino acids, and plasmids used to produce SSB mutants.

ncAA	Synthetase/tRNA (plasmid)	References	Sites selected for incorporation
TMSK	pRSF-Ch PylTMSKRS	[4]	Q16/P24
TMSNK	pRSF-G1 TMSNKRS	[12]	P24
TFAK	pRSF-G1 TFAKRS	[8]	Q16
mCF_3Phe	pRSF-G1 *m*CNP34RS[a]	[28]	F60
4FTrp[b]	pRSF-G1 F4W27RS	[10]	W54
5FTrp[b]	pRSF-G1 F4W27RS	[10]	W54
6FTrp[b]	pRSF-G1 F6W53RS	[10]	W54
7FTrp[b]	pRSF-G1 (7F-Trp) RS	[9]	W54
2,6-diFPhe	pRSF-G1 2,6-diFPheRS	[11]	F60
3,5-diFPhe	pRSF-G1 3,5-diFPheRS	[11]	F60
2,6-diFTyr	pBK-G1 2,6-diFTyrRS	This article	Y22
3,5-diFTyr	pBK-G1 3,5-diFTyrRS	This article	Y22

[a]This aminoacyl-tRNA synthetase is identical to the RS enzyme identified previously for *m*-cyanopyridylalanine. The activity of this synthetase with *m*CF_3-Phe was identified in a screen of polyspecificity; [b]The FTrp versions were produced by adding the respective fluoroindole to the bacterial growth medium, relying on the activity of native TrpB to produce the amino acids [59].

relative to the wild-type amino acid sequence. Table 1 reports the names of the RS enzymes and plasmids used and the sites selected in SSB.

Protein expression, purity, and stability

To maximize readthrough of the amber stop codon and avoid truncation, the SSB mutants were expressed in *E. coli* B-95.ΔA cells, which are devoid of the release factor RF1 [26]. Intact protein mass spectrometry indicated that glutamine was occasionally incorporated in response to the amber stop codon (Figs. S1b, S2, and S3e,h). While intact protein mass spectrometry cannot distinguish the masses of glutamine and lysine, we nonetheless refer to the effect as glutamine suppression as the glutamine codon CAG is similar to the amber stop codon TAG. In our experience, glutamine suppression is common for less efficient RS enzymes. Table 2 shows that the protein yields and purities depended on the substitution site (as observed, e.g., for TMSK in positions 16 and 24) and the amino acid (as observed most clearly for position 60).

All samples were prepared for NMR measurements in a buffer containing 300 mM NaCl and 20 mM phosphate, pH 7.2, with an initial protein concentration of 10 μM SSB tetramer (12.5 μM in the case of samples with 5FTrp and 7FTrp). During prolonged measurement times at 25 °C as well as during storage in frozen conditions, some of the SSB mutants proved to be prone to degradation and precipitation. This combined with limited spectrometer availability resulted in poor signal-to-noise ratios for some of the protein mutants that were prone to precipitation.

Table 2. Summary of the purities and yields of the SSB mutants obtained.[a]

SSB mutant	Amount of ncAA added to cell culture (mM)	Yield ($mg \cdot L^{-1}$)	Glutamine suppression (%)
Q16/TMSK	2[b]	15	0
P24/TMSK	2[b]	4.3	20
	3[b]	3.9	0
P24/TMSNK	2[b]	12	45
	4[b]	10.1	25
Q16/TFAK	20[b,c]	13	0
F60/*m*CF_3Phe	2[d]	11.1	0
W54/4FTrp	0.5[d]	4.4	0
W54/5FTrp	1[d]	2.6	0
W54/6FTrp	2[d]	1	25
W54/7FTrp	2[d]	1.2	10
F60/2,6-diFPhe	2[d]	1.1	0
F60/3,5-diFPhe	2[d]	0.6	30
Y22/2,6-diFTyr	2[d]	3.3	0
Y22/3,5-diFTyr	2[d]	2.7	0

[a]Yields refer to purified proteins. NMR measurement times are reported only for constructs where NMR spectra were recorded successfully; [b]Protein expression in Terrific Broth (TB) medium; [c]TFAK is inexpensive and nontoxic. The high concentration helps to prevent the incorporation of glutamine; [d]Protein expression in Luria-Bertani (LB) medium.

Glutamine suppression associated with TMSK mutants

TMSK was introduced into SSB to replace either Q16 or P24. The mass spectrum of the mutant SSB Q16/TMSK showed no sign of glutamine suppression (Fig. S1a), whereas the mass spectrum of the mutant

SSB P24/TMSK indicated that about 20% of the protein contained glutamine instead of TMSK, when TMSK was provided at a concentration of 2 mM (Fig. S1b). This problem of the P24/TMSK mutant was remedied by providing 3 mM TMSK (Fig. S1c). This shows that the ease of amber suppression depends not only on the intrinsic efficiency of the RS enzyme but also on the codon context.

SSB Q16/TMSK

Unexpectedly, the ^{1}H-NMR spectrum of SSB Q16/TMSK showed three distinct peaks near 0 ppm (−0.28, −0.45, and − 0.64 ppm) with an intensity ratio of approximately 1 : 3 : 1 (Fig. 3). Following titration with dT_{66}, the signals collapsed into a single resonance. This indicates that the TMS group finds itself in at least three different environments in the different monomers of the apo-protein and that these different environments do not interconvert sufficiently rapidly to average the ^{1}H-NMR signals. The effect was reproduced in independent sample preparations.

Titration of SSB Q16/TMSK with dT_{66} in increasing molar ratios up to 1 : 1 (ratios reported as ssDNA to SSB tetramer) resulted in the appearance of a new peak at −0.25 ppm, which was detectable already at the lowest titration ratio (Fig. 3A). The co-existence of the signals of free SSB and complex with ssDNA indicates slow exchange on the NMR time scale (100 Hz). Upon reaching a 1 : 1 stoichiometric ratio, the new peak became the sole TMS resonance, suggesting that in the fully bound complex, each TMS group in the SSB tetramer interacts with ssDNA in a similar way, as expected for the $(SSB)_{65}$ binding mode. Remarkably, the resonance featured a full linewidth at half-maximum of only 23 Hz after Fourier transformation without prior window multiplication.

To assess the sensitivity of the TMSK probe to different nucleotide types, we also probed SSB Q16/TMSK in a 1 : 1 complex with a different ssDNA construct, $dT_{1–36}C_{37–66}$. For this ssDNA, the $(SSB)_{65}$ model predicts that two of the SSB monomers and their associated TMS groups interact with thymidine nucleotides, while the other two interact with cytidines. Indeed, two peaks of similar intensity were observed (Fig. 3B), with one of the peaks appearing at the same chemical shift as the resonance observed with dT_{66}. This result shows that TMSK in position Q16 is sensitive to the type of nucleotide nearby.

In an attempt to resolve the proximity of a shorter segment of cytidines to a single SSB unit in the tetramer, a 1 : 1 complex was also produced of SSB Q16/TMSK with $dT_{1–50}C_{51–66}$. In this case, two peaks were observed with an intensity ratio of approximately 3 : 1 (Fig. 3B), but the minor peak was only visible as a shoulder to the main thymidine-associated resonance and its chemical shift was different from the complex with $dT_{1–36}C_{37–66}$. This suggests that the SSB subunits interact with thymidine in a conserved manner as in the complex with dT_{66} but the interaction with the poly-dC segment is more dynamic.

SSB P24/TMSK

The expression yield of SSB P24/TMSK was almost four times lower than that of SSB Q16/TMSK, which limited the signal-to-noise ratio in the ^{1}H-NMR spectra.

The 1D ^{1}H-NMR spectrum of free SSB P24/TMSK displayed a single resonance at −0.02 ppm, suggesting a uniform chemical environment for all TMS groups within the tetramer (Fig. 4). The addition of an equimolar amount of dT_{66} produced a somewhat broader resonance at a new chemical shift, indicating that the site senses the binding of ssDNA. Titration with $dT_{1–36}C_{37–66}$ and $dT_{1–50}C_{51–66}$ resulted in overlapping peaks, one at the site characteristic of nearby thymidine and the other slightly upfield, similar to the case

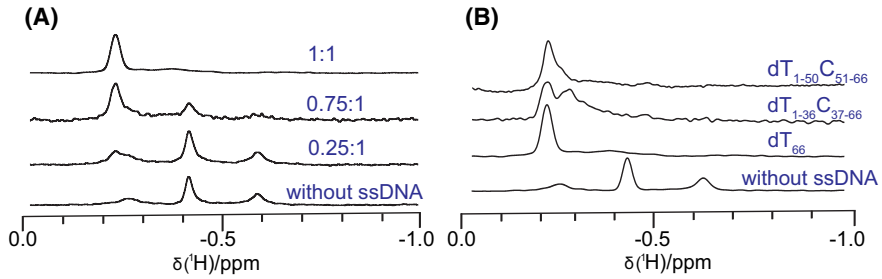

Fig. 3. The ^{1}H-NMR resonances of the TMS groups in SSB Q16/TMSK. Data processed with Lorentz–Gauss window multiplication (GB = 0.2, LB = −15 Hz). (A) Titration with dT_{66}. The molar ratios are indicated. Spectrum of the apo-protein measured in 20 min. (B) ^{1}H-NMR resonances of the TMS groups without DNA, in the presence of dT_{66}, $dT_{1–36}C_{37–66}$, and $dT_{1–50}C_{51–66}$.

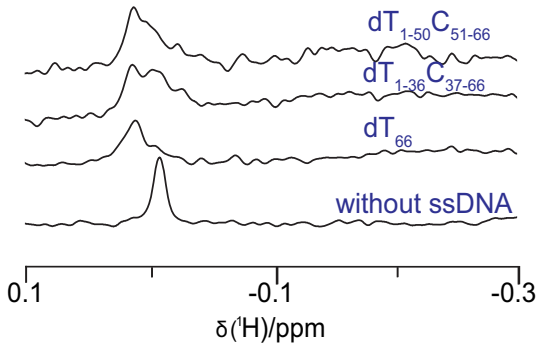

Fig. 4. ^{1}H-NMR spectra of the TMS groups in the tetramer of SSB P24/TMSK before adding ssDNA and in the presence of dT_{66}, $dT_{1–36}C_{37–66}$ and $dT_{1–50}C_{51–66}$. Data processed with window multiplication by a cosine function for improved signal-to-noise ratio. The spectrum of the apo-protein was measured in 20 min.

of SSB Q16/TMSK. SSB samples produced with TMSK proved chemically unstable as evidenced by the gradual appearance of a new TMS signal in the NMR spectrum at 0.23 ppm, which corresponds to the chemical shift of trimethylsilanol. The degradation occurred even after one week of storage at −80 °C.

SSB P24/TMSNK

Assuming that TMSK was unstable due to hydrolysis of its ester group, we installed the more stable amide analogue TMSNK to replace P24. Protein expression in the presence of 2 mM TMSNK produced protein with a yield almost threefold greater than that obtained for the corresponding TMSK mutant, but the mass spectrometric analysis indicated about 45% glutamine suppression (Fig. S2a).

Despite the chemical heterogeneity, the 1D ^{1}H-NMR spectrum of the SSB P24/TMSNK tetramer showed a single peak, suggesting that the P24 sites in the tetramer do not sense each other's chemical identity. As in the case of SSB P24/TMSK, the titration with dT_{66} changed the chemical shift downfield but the line width was narrower (8 versus 14 Hz). Titration with $dT_{1–36}C_{37–66}$ yielded a peak with a shoulder but the difference in environment posed by the segments of thymidines and cytidines produced only poorly resolved peaks. Unexpectedly, the complex with $dT_{1–50}C_{51–66}$ showed two peaks of equal intensity (Fig. 5A).

To produce SSB P24/TMSNK without glutamine suppression, protein samples were expressed also in the presence of 4 mM TMSNK. This reduced the glutamine suppression to about 25% but did not eliminate it (Fig. S2b). Repeating the titration with different ssDNA constructs for this sample, the addition of dT_{66} produced a single peak as previously but the complexes with $dT_{1–36}C_{37–66}$ and $dT_{1–50}C_{51–66}$ displayed two overlapping peaks with variable intensity ratios (Fig. 5B). We speculate that the second peak appearing when the ssDNA contains cytidine is an artifact generated by the presence of monomers containing glutamine instead of TMSNK. As the TMS signals did not fully and unambiguously resolve the different environments produced by segments of thymidine or cytidine, we explored the use of ^{19}F labels.

SSB Q16/TFAK

Aiming for the high sensitivity of a CF_3 group at the end of a long and flexible amino acid side chain, TFAK was installed at the site of Q16. The expression yield of SSB Q16/TFAK was excellent (Table 2) and glutamine suppression was negligible (Fig. S3a).

The ^{19}F-NMR spectrum recorded of SSB Q16/TFAK showed a single peak at −75.55 ppm, indicating that, unlike SSB Q16/TMSK (Fig. 3), the CF_3 groups experienced the same chemical environment in all monomers of the SSB tetramer. The addition of dT_{66} to the sample caused a clear shift of the TFA resonance. The titration with an equimolar quantity of $dT_{1–36}C_{37–66}$ produced two peaks as expected when two of the TFA groups are interacting with thymidines and the other two with cytidine (Fig. 6). Baseline resolution between the two peaks, however, was not achieved.

SSB F60/*m*CF_3Phe

The crystal structures of SSB indicate that F60 is partially solvent exposed in the free protein and close to ssDNA in the cocrystal structure, with the ssDNA stacking against the side chains of F60 and W40 [19]. This suggests that F60 could serve as a sensitive probe of the binding of different ssDNA constructs but that fluorine atoms in the *para*-position of the aromatic ring could clash with the ssDNA and thus interfere with the ssDNA-binding mode. Therefore, we substituted F60 by *m*CF_3Phe, which reduces the chance of steric clashes by allowing the CF_3 group to be positioned on either side of the phenyl ring. A previously published polyspecific aminoacyl-tRNA synthetase has been shown to install *m*CF_3Phe in low yields [27]. Fortunately, an RS variant recently identified by us for the genetic encoding of *m*-cyanopyridylalanine [28] proved to be highly active also with *m*CF_3Phe. The replacement of F60 by *m*CF_3Phe produced protein

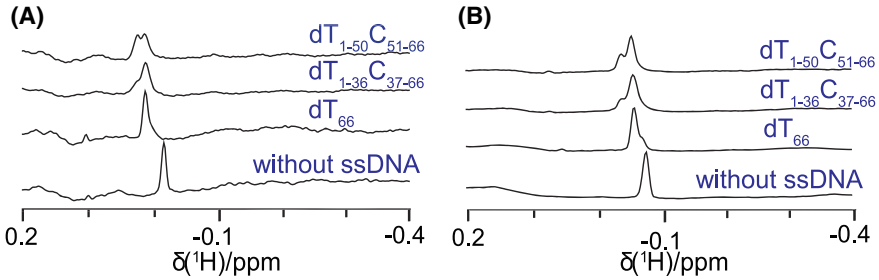

Fig. 5. 1D ^{1}H-NMR spectra of SSB P24/TMSNK. The spectra show the TMS resonance in the absence of ssDNA and in complexes made with one equivalent of dT_{66}, $dT_{1–36}C_{37–66}$, or $dT_{1–50}C_{51–66}$, respectively. Data processed with Lorentz–Gauss window multiplication (GB = 0.2, LB = –15 Hz). (A) Protein produced in the presence of 2 mM TMSNK, leading to 45% glutamine suppression. Spectrum of the apo-protein measured in 6.6 h. (B) Protein produced in the presence of 4 mM TMSNK, resulting in 25% glutamine suppression. Spectrum of the apo-protein measured in 19.7 h.

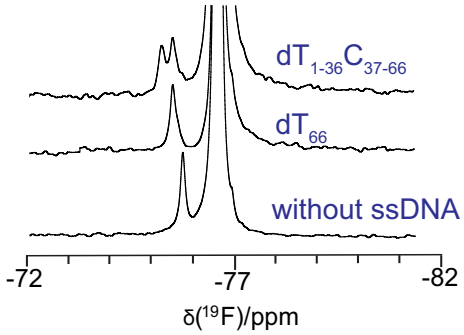

Fig. 6. 1D ^{19}F-NMR spectrum of the SSB Q16/TFAK tetramer in the apo-protein and in the presence of equimolar amounts of dT_{66} or $dT_{1\text{-}36}C_{37\text{-}66}$. The protein concentration was 10 μM in SSB tetramer. Trifluoroethanol (TFE) was added as an internal reference in large excess (0.5 mM) to render the TFE chemical shift independent from protein–TFE interactions, yielding the intense peak at −76.45 ppm. Data processed with exponential window multiplication (LB = 40 Hz). The spectrum of the apo-protein was measured in 1 h.

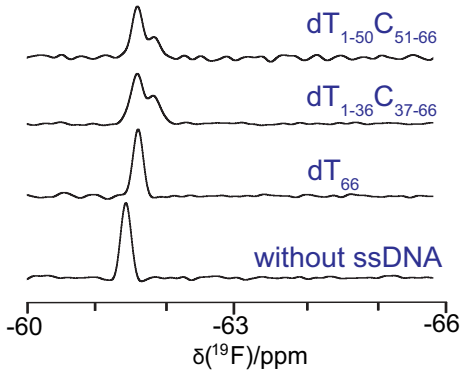

Fig. 7. 1D ^{19}F-NMR spectra of SSB F60/mCF$_3$Phe before and after adding dT_{66}, $dT_{1–36}C_{37–66}$ or $dT_{1–50}C_{51–66}$. The chemical shifts are referenced to internal TFA at −75.25 ppm. Data processed with Lorentz–Gauss window multiplication (GB = 0.2, LB = −100 Hz). Spectrum of the apo-protein measured in 44 min.

in good yield with minimal glutamine suppression (Fig. S3b).

The 1D ^{19}F-NMR spectrum of free SSB F60/mCF$_3$Phe showed a single peak at −61.46 ppm (Fig. 7). Addition of dT_{66} resulted in a small upfield shift. The complexes with $dT_{1–36}C_{37–66}$ and $dT_{1–50}C_{51–66}$ displayed a new peak slightly upfield. The intensity ratios suggest that the different peaks indicate the proximity of different SSB units to either thymidine or cytidine, assuming that the peak reporting proximity to cytidine is somewhat broader than the peak sensing thymidine (Fig. 7).

SSB W54/FTrp mutants

There are four tryptophan residues in the SSB monomer, all of which are solvent exposed. Selecting W54 for substitution by either 4FTrp, 5FTrp, 6FTrp, or 7FTrp, we explored the potential of these ncAAs as ^{19}F-NMR sensors of ssDNA binding. In these ncAAs, the fluorine atom is part of a relatively rigid amino acid side chain (Fig. 1B), and 4FTrp and 5FTrp are not expected to engage in direct fluorine–ssDNA contacts.

The expression yields of SSB W54/4FTrp, SSB W54/5FTrp, SSB W54/6FTrp, and SSB W54/7FTrp were acceptable (Table 2), but intact protein mass spectrometry indicated variable success in installing the fluorinated tryptophan analogues site-selectively and without glutamine suppression (Fig. S3c–f). In the case of SSB W54/4FTrp, nearly 20% of the SSB protein contained two FTrp residues, suggesting unintentional substitution of native Trp residues by 4FTrp (on average about 7% at each site, Fig. S3c). In contrast, SSB W54/5FTrp appeared homogeneous with

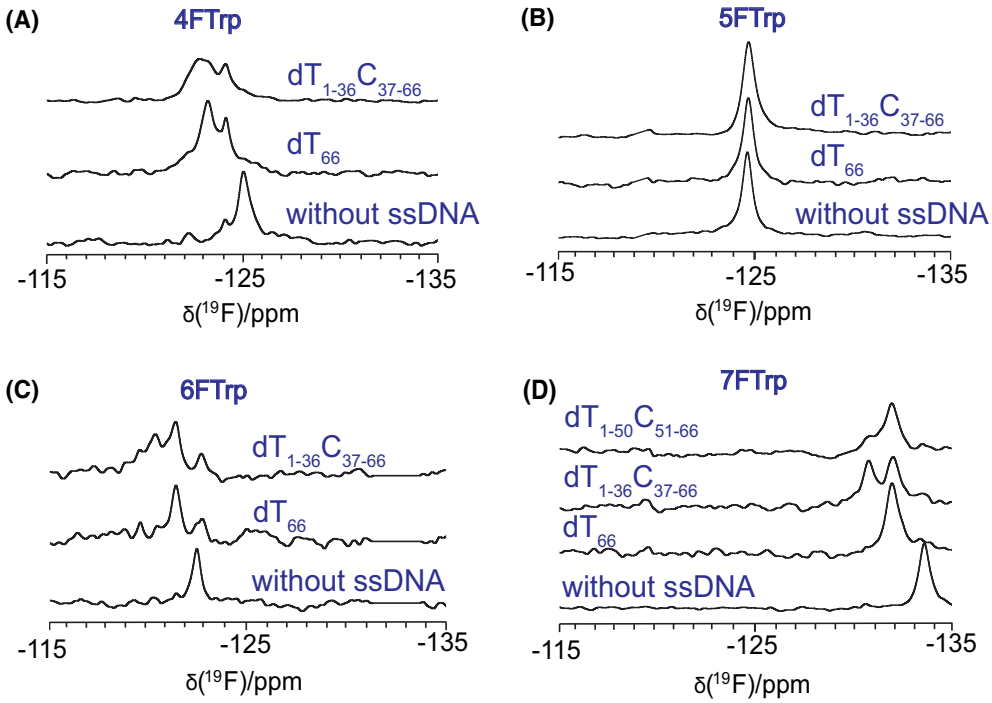

Fig. 8. 1D ^{19}F-NMR spectra of the SSB W54/FTrp mutants before adding ssDNA, with dT_{66}, with $dT_{1–36}C_{37–66}$ and with $dT_{1–50}C_{51–66}$. Data processed with exponential window multiplication (LB = 100 Hz). (A) Spectra recorded of SSB W54/4FTrp. The recording time of the apo-protein was 1.4 h. (B) SSB W54/5FTrp. Recording time of the apo-protein was 2.8 h. (C) SSB W54/6FTrp. Recording time of the apo-protein was 16 h. (D) SSB W54/7FTrp. Recording time of the apo-protein was 15 h.

full substitution of W54 by 5FTrp (Fig. S3d) but SSB W54/6FTrp and SSB W54/7FTrp showed evidence of 25% and 10% glutamine suppression, respectively (Fig. S3e,f). These effects correlated with the protein expression yields, where 4FTrp yielded more protein than 5FTrp, and 6FTrp and 7FTrp yielded the smallest amount of protein (Table 2). 4FTrp and 5FTrp were installed using the same polyspecific RS enzyme in the same cell line with the same plasmids.

The SSB construct used in the present work was prone to precipitation at tetramer concentrations above 4 μM under the conditions used for NMR (300 mM NaCl, 20 mM phosphate pH 7.2, 25 °C). The incorporation of a ncAA notably modulated the rate of precipitation. In the case of the W54/FTrp mutants, all precipitated particularly quickly, which limited the time available for NMR measurements. The mutant with 6FTrp was most sensitive to precipitation while the mutant with 7FTrp was least sensitive. The mutant with 5FTrp became insoluble even during storage at −80 °C.

In the absence of ssDNA and despite the chemical heterogeneities arising from glutamine suppression, the ^{19}F-NMR spectra showed single peaks as expected for a symmetric SSB tetramer (Fig. 8). In the presence of ssDNA, the spectrum of the W54/5FTrp mutant changed very little, except for an increase in line width from about 200 to 230 Hz with dT_{66} and 300 Hz with $dT_{1–36}C_{37–66}$ (Fig. 8B). The W54/7FTrp mutant showed a change in chemical shift in the presence of dT_{66}, a split of the signal into two with $dT_{1–36}C_{37–66}$ and a weaker second peak with $dT_{1–50}C_{51–66}$ (Fig. 8D). This pattern follows the expectation that position 54 in each monomer senses a segment made of either thymidines or cytidines.

The spectra of the W54/4FTrp and W54/6FTrp mutants were more difficult to interpret, either because a new, relatively narrow signal at −124.2 ppm emerged over time and the peaks arising from polythymidine and polycytidine were difficult to resolve (Fig. 8A) or because the signal-to-noise ratio was limiting and the sample appeared heterogeneous in the presence of ssDNA (Fig. 8C). Although most of the fluorotryptophan isomers responded to the presence of ssDNA with chemical shift changes of multiple ppms, the widths of the ^{19}F-NMR signals were also large, thus preventing baseline resolution between different signals and compromising the height and sensitivity of the NMR peaks.

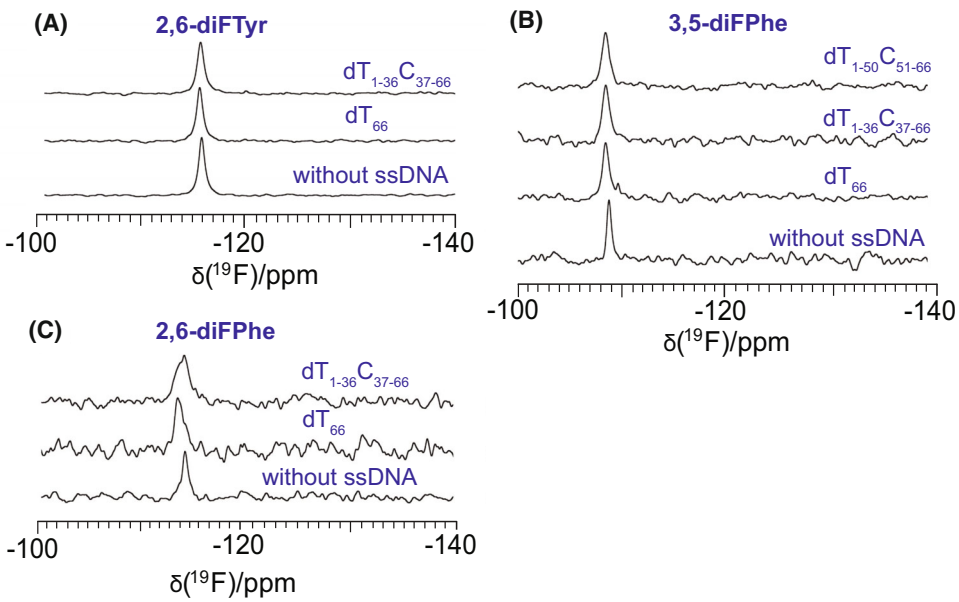

Fig. 9. ^{19}F-NMR spectra of SSB Y22 and SSB F60 mutants in the free state and in complexes with dT_{66}, $dT_{1–36}C_{37–66}$ and $dT_{1–50}C_{51–66}$. The data were processed with exponential window multiplication, using a line broadening value of 200 Hz. (A) SSB Y22/2,6-diFTyr. The spectrum of the apo-protein was recorded in 19 h. (B) SSB F60/3,5-diFPhe. The spectrum of the apo-protein was recorded in 4.4 h. Due to precipitation, samples with ssDNA were made using fresh protein. (C) SSB F60/2,6-diFPhe. The apo-protein precipitated significantly during the recording time (20 h). Samples of the complex with ssDNA were made using fresh protein.

Genetic encoding of 2,6-diFTyr and 3,5-diFTyr

Assuming that the enhanced sensitivity of FTrp mutants towards precipitation arises from the increased hydrophobicity of CF versus CH groups, we turned to the more polar ncAAs 2,6-diFTyr and 3,5-diFTyr. An aminoacyl-tRNA synthetase specific for these two amino acids was selected from an established library of RS enzymes derived from the pyrrolysyl-tRNA synthetase of the methanogenic archaeon ISO4-G1 (G1PylRS), which has also been the basis for previously published genetic encoding systems of fluorotryptophan isomers and 7-azatryptophan [9,10,29]. Briefly, *E. coli* DH10B cells were cotransformed with the library plasmid pBK-G1PylRS and the selection plasmid pBAD-H6RFP encoding the gene of a red fluorescent protein (RFP) interrupted by an amber stop codon. Fluorescence-activated cell sorting was used to select RFP-positive cells grown in the presence of 1 mM ncAA and RFP-negative cells grown without ncAA. Multiple selection rounds progressively enriched cells expressing both active and specific G1PylRS variants. About 60 colonies expressing high levels of RFP were grown from the final cell culture for each ncAA, isolated and sequenced. The best mutants were G1-2,6-diFTyrRS (carrying the mutations L124C, Y125M, N165T, V167F, Y204W, A221S, and W237S) and G1-3,5-diFTyrRS (carrying the mutations N165A and W237R; Table S1).

Difluorotyrosine mutants of SSB

The ncAAs 2,6-diFTyr and 3,5-diFTyr were used to replace Y22 and test as probes of nearby thymidines and cytidines. The mass spectrometric analysis of SSB Y22/2,6-diFTyr confirmed the complete and selective incorporation of the ncAA (Fig. S3g) but the expression yield was low (Table 2). The ^{19}F-NMR spectrum of the protein showed a single peak which, relative to the line width, shifted only a little in response to dT_{66} or $dT_{1–36}C_{37–66}$ (Fig. 9A). Most notably, the formation of the complexes was associated with an increase in line width (from about 230 Hz in the free protein to about 250 Hz in the presence of polythymidine and 290 Hz in the presence of $dT_{1–36}C_{37–66}$). While the natural widths of the ^{19}F-NMR signals made it difficult to discern more obvious spectral changes, closer proximity to the ssDNA may well have led to larger effects. The crystal structure indicates that the fluorine atoms are more than 10 Å from the ssDNA, making direct contacts unlikely. Notably, the fluorine atoms in 2,6-diFTyr decrease the pK_a value of the hydroxyl group to about 8.1 [30], raising the possibility that the ^{19}F-

NMR spectra sense a change in the protonation equilibrium of the hydroxyl group caused by the presence of the polyanionic ssDNA.

Assuming that this effect would be even greater in 3,5-diFTyr, where the hydroxyl group has a pK_a value closer to neutral (6.8 [30]), we produced SSB Y22/3,5-diFTyr. Unfortunately, the yields were too low for NMR spectroscopy (Table 2). Furthermore, the mass spectrometric analysis indicated almost 30% glutamine suppression (Fig. S3h).

Difluorophenylalanine mutants of SSB

To position the fluorine atoms closer to the ssDNA while avoiding possible long-range effects arising from pK_a changes, we produced the mutants SSB F60/2,6-diFPhe and SSB F60/3,5-diFPhe. The first mutant was obtained in good yield (Table 2) and the mass spectrometric analysis demonstrated complete and selective incorporation of the ncAA (Fig. S3i,j). The ^{19}F-NMR spectra of both mutants showed a single peak (Figs 9B,C), but only a small change in chemical shift after the addition of ssDNA. In the case of SSB F60/2,6-diFPhe, the formation of complex was indicated by an increase in line width from about 170 Hz in the free protein to about 440 Hz in the presence of dT_{66}. In the case of SSB F60/3,5-diFPhe, the line width increased from about 83 Hz to 230 Hz in the presence of dT_{66}. With ssDNA containing segments of polycytidine, the peaks broadened further, suggesting that contacts with cytidines produce a different chemical shift. In the case of SSB F60/2,6-diFPhe, the line shape became distinctly heterogeneous. Based on the crystal structure, the fluorine atoms of 3,5-diFPhe are expected to be closer to the ssDNA than those of 2,6-diFPhe. As in the case of the fluorinated tryptophan and tyrosine analogues, however, the intrinsic line widths were too great to resolve the different chemical environments generated by the different nucleotide sequences, even when the chemical shift changes were greater than 0.5 ppm.

Overview

Table 3 summarizes the magnitudes of chemical shift changes and line widths observed in *E. coli* SSB with the different ncAAs investigated. Given the low protein concentrations (10 μM in SSB tetramer) combined with the use of 3 mm NMR tubes, it is encouraging that the ^{19}F-NMR signals of ncAAs containing single-fluorine atoms can be detected in this high-molecular weight system. The fluorinated aromatic ncAAs produced the largest chemical shift changes, but none of them displayed a line width below 100 Hz in the complex with dT_{66}, thereby compromising the spectral resolution.

Table 3. Chemical shift differences of different NMR probes observed between free and dT_{66}-bound *E. coli* SSB mutants.

SSB mutant	Chemical shift change (ppm)	Line width (Hz) in apo-protein	Line width (Hz) in complex with dT_{66}
SSB Q16/TMSK	0.20	42, 19, 30[a]	23
SSB P24/TMSK	0.04	8	14
SSB P24/TMSNK	0.02	8	9
SSB Q16/TFAK	0.26	32	44
SSB F60/*m*CF_3Phe	0.18	64	69
SSB W54/4FTrp	1.84	250	730
SSB W54/5FTrp	0.06	150	180
SSB W54/6FTrp	1.10	90	110
SSB W54/7FTrp	1.72	200	310
SSB Y22/2,6-diFTyr	0.20	180	200
SSB F60/2,6-diFPhe	0.70	170	440
SSB F60/3,5-diFPhe	0.30	83	230

[a]Line widths of the three signals observed (low-field to high-field).

The TMS and TFA groups of the aliphatic ncAAs are at the ends of long and flexible side chains. Their NMR signals are relatively narrow as they reorientate much faster than the SSB tetramer. Given their high solvent exposure both in the free protein and the complex with ssDNA, their chemical shifts were changed relatively little by the presence of ssDNA, resulting in an effective spectral resolution comparable with that of the fluorinated aromatic ncAAs positioned much closer to the ssDNA. Their narrow line widths, however, lifted the NMR sensitivity.

Discussion

The aim of the present project was to investigate the binding of ssDNA to *E. coli* SSB in the $(SSB)_{65}$ mode in aqueous solution by labelling different sites with ncAAs that would minimally perturb the structure of the protein or its interaction with ssDNA (Fig. 1B). In addition, we expected the ncAAs to respond differently to nearby segments of poly-dT and poly-dC. All ncAAs investigated proved suitable for selective detection by NMR spectroscopy without isotopic labelling despite the high-molecular weight of the complex. Most of them responded differently to different nucleotide segments in the ssDNA.

^{19}F ncAA probes

The advantage of ^{19}F-modified amino acids as background-free NMR probes of high-molecular

weight systems has long been recognized [31]. To gain site-specific information without spectral overlap, however, the probes must be installed selectively at a single site. The most general approach uses chemical modifications of single-cysteine mutants. Numerous chemical tags have been developed to introduce ^{19}F-labels in this way [32–35]. Generating single-cysteine mutants, however, often requires extensive protein engineering to remove natural cysteine residues from the wild-type protein. In addition, the tagged cysteine residue is dissimilar in structure to any canonical amino acid and the posttranslational chemical labelling requires sample handling, which is undesirable for unstable proteins such as SSB.

In contrast, *in vivo* genetic encoding of ncAAs allows facile installation of the NMR probe in a cost-effective way independent of the amino acid composition. In addition, the genetically encoded ncAA can be very similar to the canonical amino acid it replaces. Specifically, single-fluorine analogues of the aromatic amino acids phenylalanine, tyrosine and tryptophan have been successfully encoded genetically [9–11,36–38]. For enhanced sensitivity and selectivity, genetic encoding of ncAAs with CF_3 groups has proven particularly useful [6,39,40].

^{19}F chemical shifts sensitively report on local changes in van der Waals interactions and electric field [25,41], but hardly on long-range electrostatic effects [42]. Binding of a polyanion like ssDNA would thus be expected to report primarily on local effects, especially in the presence of a high concentration of salt. If long-range effects had a significant impact, more complex spectra would have resulted as ssDNA binds and breaks the four-fold symmetry of the SSB tetramer.

Electric field effects have been predicted to contribute on the order of 10 ppm to ^{19}F chemical shifts [43]. Such a strong predominant influence of electrostatic effects on the ^{19}F chemical shifts was not apparent in the present work. Specifically in the cases of *m*CF_3Phe, TFAK, and 4FTrp, formation of the complex with dT_{66} did not result in much greater chemical shift changes than the additional change elicited by segments of cytidines.

In general, CF_3 groups respond to ligand binding with smaller chemical shift changes than single ^{19}F spins [5,6,31,44–46]. In the case of SSB, 3,5-diFTyr installed remote from the ssDNA-binding site to minimize structural perturbations resulted in minimal chemical shift changes. Also, the mutant F60/*m*CF_3Phe produced relatively small chemical shift changes considering its very close proximity to the ssDNA (Fig. 1B).

It has recently been shown that the sensitivity of ^{19}F chemical shifts towards changes in the environment can be enhanced by labelling with CH_2F groups, which respond sensitively to small conformational changes due to a γ-gauche effect [47,48], or with a CF_3-tagged 2-pyridone that is sensitized by tautomeric effects [49]. Genetic encoding of ncAAs containing these moieties poses an attractive future direction.

Aliphatic ncAAs

The ncAAs with a TMS group at the end of a flexible side chain produced the narrowest NMR signals but could be installed only remote from the ssDNA-binding site to avoid interfering with the complex formation. The ^{1}H chemical shifts of TMSK and TMSNK responded sensitively to the binding of ssDNA but were less sensitive to poly-dC versus poly-dT segments. Despite the shorter side chain of TFAK, this aliphatic ncAA resolved the differences between apo SSB and the complexes with poly-dT and poly-dC relatively well, while avoiding glutamine suppression effects and delivering good sensitivity. In a previous comparison of TMSK, TMSNK, and TFAK in the SARS-CoV-2 main protease, the ^{19}F-NMR spectrum of TFAK likewise showed greater chemical shift changes upon ligand binding than the ^{1}H-NMR spectra of TMSK and TMSNK [12].

As opposed to fluorine-labeled tyrosine and phenylalanine analogues, the NMR signals of aliphatic ncAAs cannot be broadened by chemical exchange associated with aromatic ring flips. In the case of SSB F60/*m*CF_3Phe, the narrower ^{19}F-NMR signal in the complex with dT_{66} may be explained by better suppression of the aromatic ring flip in the complex.

Symmetry considerations for SSB

In the absence of ssDNA, all SSB mutants displayed a single resonance, consistent with a symmetric tetramer, where each constituting monomer presents the same environment to the ncAAs. The mutant SSB Q16/TMSK was the only exception, displaying three TMS signals in the ^{1}H NMR spectrum. A possible explanation could be fortuitous binding of the TMS group to unidentified hydrophobic sites in the protein. In principle, interactions could also occur between different SSB tetramers, but the high-molecular weight of an octamer would disfavor their observation by NMR. The phenomenon was not recapitulated by SSB Q16/TMSNK or SSB Q16/TFAK.

In the system of tetrameric SSB with ssDNA, heterogeneous and broader lines may arise for all ncAA

probes as the four-fold symmetry of the SSB tetramer is broken in the complex with ssDNA. Indeed, the SSB mutants P24/TMSK, P24/TMSNK, W54/4FTrp, and W54/6FTrp showed evidence for inhomogeneous signals in the presence of dT_{66} (Figs 4, 5, and 8), which did not correlate with the degree of glutamine suppression (Table 2). In contrast, the other mutants resolved no heterogeneities in the complex with dT_{66} as expected for a very similar binding mode of the ssDNA to each of the SSB monomers. Another potential mechanism of line broadening arises from chemical exchange between bound and unbound ssDNA segments although the binding affinity of poly-dT is known to be extremely high [16].

In the absence of ssDNA, the ssDNA-binding OB domain of SSB has been proposed to interact with the flexible C-terminal domain [17], which could break the symmetry of the tetramer. This interaction has been observed at low pH in monomeric SSB [18] but it is weak and most of the C-domain is highly flexible at neutral pH even at low salt concentration [14]. At high-salt concentrations, the C-terminal domain is in an intrinsically unstructured, highly flexible conformation, demonstrating the importance of electrostatics in its binding interactions with the OB domain [14]. To facilitate protein purification, the SSB construct used in the present work contained a C-terminal His_6-tag preceded by the TEV recognition site, which following cleavage with TEV protease, added the peptide ENLYFQ relative to the wild-type amino acid sequence. We found no evidence that the modified C terminus promoted any interactions of the C-terminal domain with the OB domain.

Sample precipitation

Our construct of wild-type SSB was prone to precipitation in high-salt buffer conditions (300 mM NaCl) at 25 °C. Several of the mutants made with ncAAs also precipitated quickly and with rates depending on the ncAA installed. Even the substitution of single hydrogen atoms by fluorine in the fluorinated analogues of aromatic amino acids proved sufficient to increase the sensitivity towards precipitation. Precipitation was observed also for the complexes with ssDNA, limiting the time available for NMR measurements. Accelerated precipitation was observed for the W54/FTrp mutants. The reason for this effect is unclear. For example, the crystal structure indicates that the fluorine atom in the W54/7FTrp mutant is solvent exposed and introduces no steric clash in apo SSB or the complex with ssDNA. Furthermore, the presence of ssDNA should hinder possible interactions of the fluorine atom with other SSB tetramers. Nonetheless, the W54/7FTrp mutant precipitated more slowly than the other W54/FTrp mutants where, according to the crystal structure of the wild-type protein, the fluorine atom would engage in van der Waals contacts with other amino acid residues.

The Y22/diFTyr mutants precipitated particularly quickly, preventing the recording of NMR spectra for SSB Y22/3,5-diFTyr. The hydrophobicity of fluorobenzene is known to be greater than that of benzene and to increase with the number of fluorine atoms [50,51]. The high solvent exposure of Y22 suggests that the increased hydrophobicity of fluorinated aromatic rings contributes decisively to protein aggregation. In agreement with the hypothesis that high solvent exposure of fluorinated aromatic residues promotes protein aggregation, the mutants F60/3,5-diFPhe and F60/2,6-diFPhe were less sensitive toward precipitation.

There is growing evidence that fluorination of canonical aromatic residues modulates the stability of proteins with mostly destabilizing effects [52], whereas similar effects have not been reported for fluorine tags installed on cysteine or lysine residues by posttranslational chemical reaction [53]. In the present work, the samples made with aliphatic ncAAs were more resilient towards precipitation, although the chemical structures of these ncAAs deviated much more from those of the canonical amino acids they replaced. Site-specific labelling in systems tolerating exclusively canonical amino acids may ultimately necessitate isotope labelling combined with photocaging [54].

Conclusion

The present work compared the effectiveness of genetically encoded ncAAs as site-specific NMR probes in a large protein–nucleic acid complex (100 kDa). While much greater spectral resolution can be achieved by 2D NMR with ^{13}C-labeled ncAAs [15,45,46], the cost of isotope labelling becomes prohibitive for unstable protein systems like SSB. In this situation, one-dimensional NMR spectra are attractive for the speed with which they can be recorded, even if the ncAA contains only a single ^{19}F spin and the protein concentration is as low as 20 μM. In principle, the chemical label of ncAAs may alter the properties of the target protein in unforeseen ways. To identify idiosyncratic effects of a specific ncAA at a specific site, it is therefore desirable to double-check by using different ncAAs. In the case of SSB, the consensus result is that all four monomers in the tetramer engage in binding to dT_{66} in the same way, as expected for the $(SSB)_{65}$ binding mode. Furthermore, segments of poly-dC can

be sensed and distinguished from poly-dT, setting the stage for more detailed investigations of sequence-dependent ssDNA binding.

Materials and methods

Materials

The amino acid TMSK was purchased from BAPEKS (Riga, Latvia). TMSNK was synthesized as described previously [12]. TFAK and the fluoroindoles were purchased from AK Scientific (Union City, CA, USA), *m*CF_3Phe from eNovation Chemicals (Green Brook, NJ, USA), 2,6-diFPhe and 3,5-diFPhe from AmBeed (Chicago, IL, USA), and 2,6-diFTyr and 3,5-diFTyr from AP Bioscience (Princeton, NJ, USA). Single-stranded DNA oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA, USA) and HPLC-purified. All ssDNAs were 66 nucleotides long. ssDNA concentrations were confirmed by absorption measurement at 260 nm, using a NanoDrop spectrophotometer at 260 nm with Milli-Q water as the blank.

Selection of aminoacyl-tRNA synthetases for 3,5-diFTyr and 2,6-diFTyr

The selection of functional G1PylRS enzymes followed a previously established protocol, employing a published library of G1PylRS mutants encoded on the pBK-G1RS plasmid [9,28]. The plasmid library was transformed into *E. coli* DH10B cells harboring the pBAD-H6RFP reporter plasmid encoding mRFP1 red fluorescent protein (RFP) with an amber stop codon following an N-terminal His_6 tag. After transformation, the culture was directly inoculated into 25 mL LB medium supplemented with 100 mg·L^{-1} carbenicillin, 50 mg·L^{-1} kanamycin, 0.4% L-arabinose, and 1 mM target ncAA. This culture served as the sample for the first round of positive selection (1P+). The same cells grown without the supplement of ncAA were used as a control sample (1P–). Overnight expression at 37 °C led to a readily detectable level of RFP expression. The cells were harvested, resuspended in 5 mL of PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and diluted 20-fold to a concentration suitable for fluorescence-activated cell sorting (FACS). FACS was performed using a FACSAria Fusion cell sorter (BD Biosciences, USA). Cells with high red fluorescence levels were collected from the 1P+ sample (indicated by violet shades in Figs 10A and 11A) and subjected to a subsequent round of negative selection (2N–) in the absence of ncAA. Cells exhibiting low RFP expression were collected from the 2N– sample and aliquoted to inoculate media under positive (3P+, with ncAA) and negative (3P–, without ncAA) conditions. The 3P+ sample for 2,6-diFTyr demonstrated a clear response to the presence of 2,6-diFTyr, and the top 5.2% of RFP-fluorescent cells from the 3P+ sample were collected. The selection experiments for 3,5-diFTyr continued to a total of seven rounds following the same strategy to enrich the target population. Approximately the top 3% of RFP-fluorescent cells from the 5P+ samples for 3,5-diFTyr were collected.

About 2000 cells collected from each final round were recovered by plating on LB agar plates containing 100 mg·L^{-1} carbenicillin and 50 mg·L^{-1} kanamycin. Isolated colonies were analyzed in 96-well plates. For each ncAA, 60 enzyme candidates were inoculated into media under both positive (with 1 mM ncAA) and negative (without ncAA) conditions. The red fluorescence intensity was measured as an indicator of RFP expression and normalized to the OD_{600} of the cell culture using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland; Figs 10B and 11B). The DNA sequence analysis of the pBK-G1RS plasmids identified five individually different candidates to incorporate 2,6-diFTyr and a single unique mutant to incorporate 3,5-diFTyr. The amino acid mutations of these candidates are listed in Table S1. The expression plasmids of G1-2,6-diFTyrRS and G1-3,5-diFTyrRS are available from the Addgene plasmid repository (Watertown, MA, USA) with the accession codes #251487 and #251488, respectively.

Plasmid preparation for protein expression

The gene encoding *E. coli* single-stranded DNA-binding protein (SSB), modified to include a C-terminal TEV protease recognition site followed by a His_6-tag (Table S2), was cloned between the *NdeI* and *XhoI* restriction sites of the vector pCDF as described previously [55,56]. The sites for installing the ncAAs were defined by an amber stop codon introduced by mutagenesis using the QuikChange method [57].

Protein expression and purification

The SSB mutants were expressed in *E. coli* B-95.ΔA cells which are devoid of the release factor RF1 [26]. The cells were cotransformed with the pCDF-SSB expression plasmid and the plasmid encoding the aminoacyl-tRNA synthetase/tRNA pair specific for each ncAA (Table 1).

Cells were cultured at 37 °C in media containing 25 μg·mL^{-1} spectinomycin and 20 μg·mL^{-1} kanamycin. TB medium was used for expressing TMSK, TMSNK, and TFAK mutants for enhanced protein yields, while LB was used to express the other mutants. The ncAAs were supplied to the media at 2 mM (TMSK, *m*CF_3Phe, diF-derivatives), 3–5 mM (TMSNK), or 20 mM (TFAK) concentration. Fluoroindole precursors were added

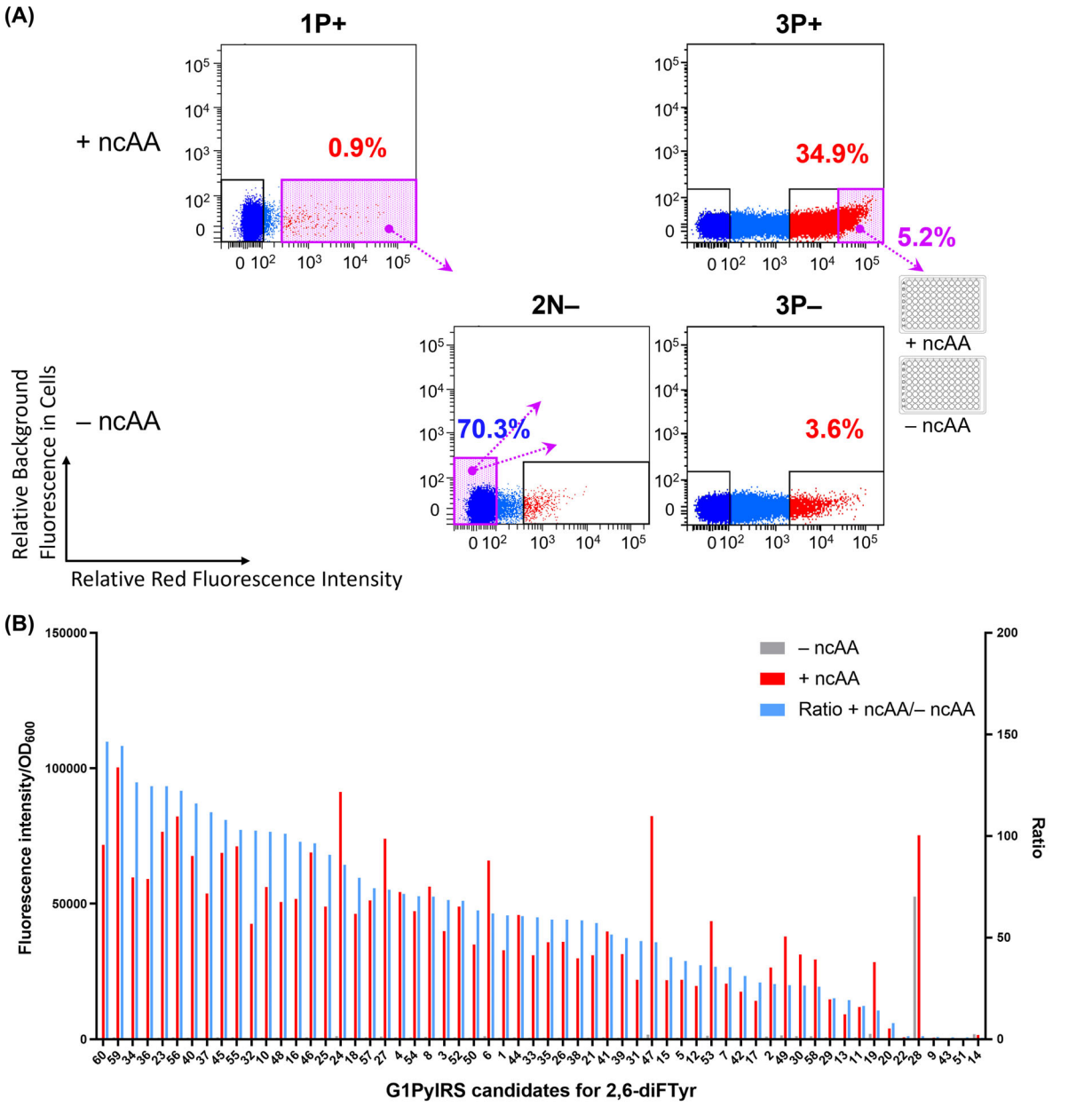

Fig. 10. FACS experiments for selecting functional G1PylRS enzymes for 2,6-diFTyr. (A) FACS screening of G1PylRS variants for activity and specificity in recognizing 2,6-diFTyr. The horizontal axis of the scatter plots represents red fluorescence intensity (excitation at 560 nm), while the vertical axis indicates background fluorescence in cells excited at 488 nm. Violet-shaded regions identify the collected cell populations. Arrows illustrate the subsequent selection strategy applied after amplification by culturing. The preparation conditions of each sample are indicated with '+', denoting growth conditions with 1 mM ncAA, and '–' for conditions without ncAA. Positive and negative selection rounds are labeled as '**P'** and '**N'**. (B) Activity and specificity screen of G1PylRS variants for 2,6-diFTyr incorporation. Cells from the 5.2% fraction with the highest red fluorescence in the final selection round were cultured in 96-well plates with and without 1 mM 2,6-diFTyr. Red fluorescence intensity indicative of the readthrough efficiency of the amber-interrupted reporter gene was then measured. The plot presents the colonies ranked in a descending order based on the ratio of red fluorescence in the + ncAA wells compared to the – ncAA wells. This ranking highlights the candidates with the highest activity and specificity for 2,6-diFTyr incorporation.

immediately before induction for fluorotryptophan incorporation in concentrations adjusted to account for their toxic effects on the *E. coli* cell culture (Table 2).

The cell cultures were induced with 1 mM IPTG when their OD_{600} value had reached 0.6–0.8 and grown overnight at room temperature. Cells were harvested (5000 × *g*,

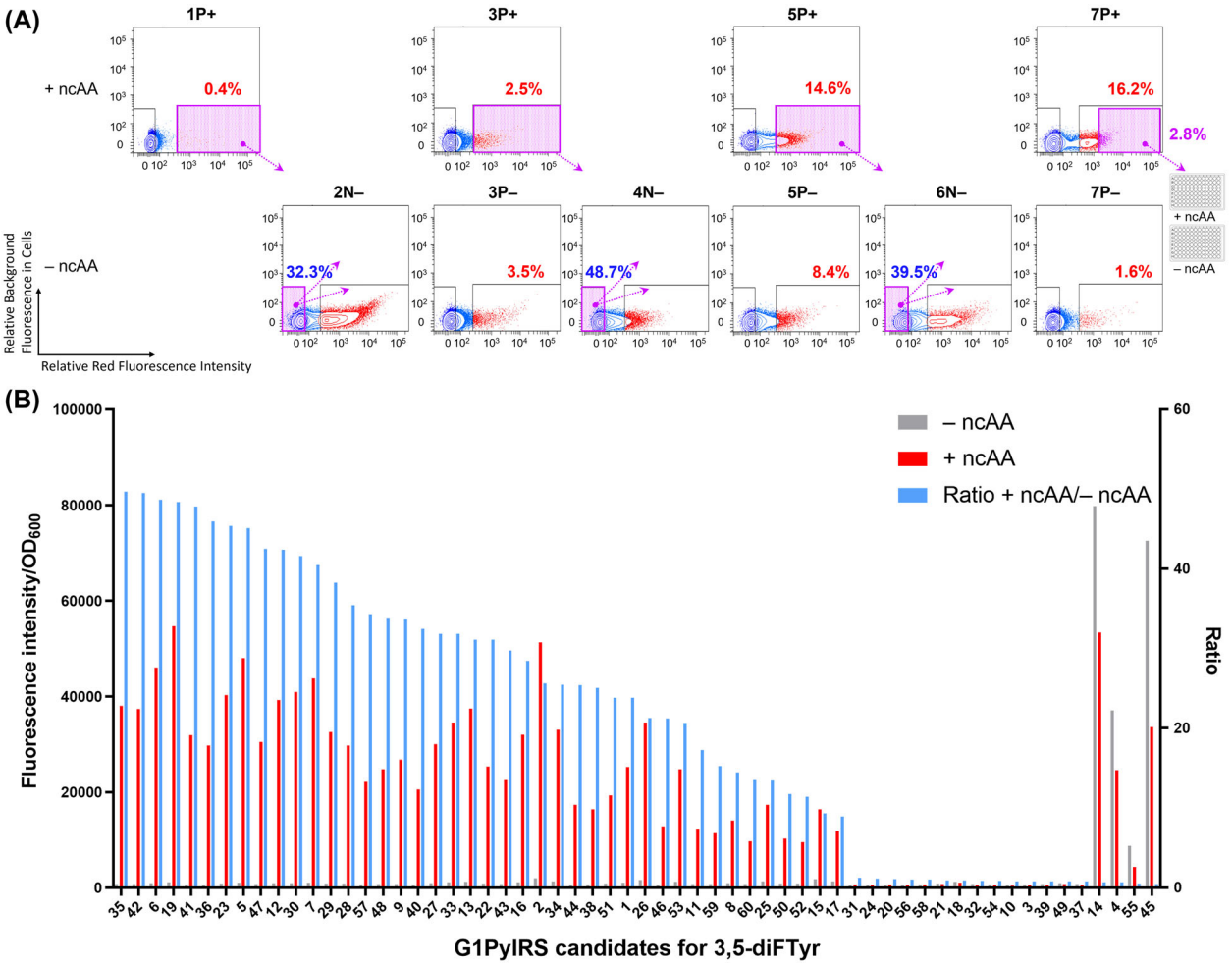

Fig. 11. FACS experiments for selecting functional G1PylRS enzymes for 3,5-diFTyr. (A) FACS screening of G1PylRS variants for activity and specificity in recognizing 3,5-diFTyr. Annotations are the same as in Fig. 10. (B) Activity and specificity screen of G1PylRS variants for 3,5-diFTyr incorporation. Cells from the 2.8% fraction with the highest red fluorescence in the final selection round were cultured in 96-well plates with and without 1 mM 3,5-diFTyr. As in Fig. 10, the plot ranks the colonies in descending order based on the ratio of red fluorescence in the + ncAA wells compared to the – ncAA wells, highlighting the candidates with the highest activity and specificity for 3,5-diFTyr incorporation.

30 min), resuspended in lysis buffer (buffer A, 50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole), and lysed using an Emulsiflex-C5 homogenizer (Avestin, Ottawa, Canada). The lysates were cleared by centrifugation (13 000 × ***g***, 1 h), filtered, and applied to a His GraviTrap column (Cytiva, Marlborough, MA, USA) equilibrated with buffer A.

Columns were washed sequentially with 20 column volumes buffer A, 20 column volumes 1 M $MgCl_2$ (prepared in water, to remove bound ssDNA) and 20 column volumes of buffer A again before elution with 10 mL of elution buffer (buffer B, same as buffer A, but with 300 mM imidazole). The purified fractions were analyzed by SDS/PAGE and intact protein mass spectrometry. The His_6-tag was removed by overnight TEV protease digestion (100 : 1 ratio, 4 °C) in buffer C (50 mM Tris/HCl, pH 8.0, 500 mM NaCl). Cleaved samples were buffer-exchanged into NMR buffer (buffer D: 20 mM phosphate buffer, 300 mM NaCl, pH 7.2), and the residual TEV protease was removed by passing the sample through a Ni-NTA column. The final construct retained a six-residue TEV cleavage remnant (ENLYFQ) at the C terminus (Table S2). Protein concentrations were determined by absorbance at 280 nm using an extinction coefficient of 29 450 M^{-1} cm^{-1}. Prior to NMR measurements, 10% D_2O was added to provide the lock signal.

NMR spectroscopy

NMR spectra were recorded using Bruker Avance NMR spectrometers operating at ^{1}H-NMR frequencies of

800 MHz for the ^{1}H-NMR spectra or 600 MHz for the ^{19}F-NMR spectra. Both spectrometers were equipped with TCI cryoprobes for 5 mm NMR tubes. All spectra were recorded at 25 °C, using 3 mm NMR tubes to account for the high-salt conditions. 1D ^{1}H-NMR spectra of SSB variants with TMSK and TMSNK were acquired using 20–40 μM protein (monomer concentration). A jump-return sequence was used to suppress the water resonance without saturating it [58]. The acquisition time was 160 ms. The recovery delay between scans was 1 s.

The ^{19}F-NMR spectra were measured without ^{1}H decoupling, using an acquisition time of 260 ms and a recovery delay between scans of 1 s. 0.1 mM trifluoroacetic acid (TFA) was added as an internal reference for the samples made with fluorinated ncAAs, except for SSB Q16/TFAK, which used 0.5 mM trifluoroethanol as the internal reference. The ssDNA was added to the SSB samples in molar ratios of ssDNA to SSB tetramer of 0.25 : 1, 0.5: 1, 0.75 : 1, and 1 : 1. For unstable protein preparations, only the 1 : 1 ratio was used. The free induction decays were processed with window functions designed to maximize the signal-to-noise ratio (mostly for the ^{19}F-NMR spectra) and spectral resolution (for the ^{1}H-NMR spectra).

Intact protein mass spectrometry

Intact protein mass spectrometric analysis was carried out using an Elite Hybrid Ion-Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA) connected to an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column (Agilent, Santa Clara, CA, USA) in an UltiMate S4 3000 UHPLC system (Thermo Scientific). About 7.5 pmol of sample was injected.

Author contributions

SMT prepared all samples involving SSB and measured the NMR spectra. HQ selected and characterized the new RS enzymes. EA cloned the final expression vectors of the new RS enzymes. JG and CN synthesized TMSNK. APW produced wild-type SSB. TH supervised HQ and the library design of RS enzymes. GO and SMT designed and coordinated the project. SMT wrote the first version of the manuscript and all authors contributed to the manuscript writing and editing.

Acknowledgements

The authors thank Dr Luke A. Adams for a sample of TMSK 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, including the Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and a Discovery Project (DP230 100079) is gratefully acknowledged. . Open access publishing facilitated by Australian National University, as part of the Wiley - Australian National University agreement via the Council of Australasian University Librarians

Conflicts of interest

The authors declare no conflicts of interest.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/febs.70446.

Data availability statement

The original data are available from the authors upon reasonable request.

References

1 Chen WN, Kuppan KV, Lee MD, Jaudzems K, Huber T & Otting G (2015) *O-tert*-butyltyrosine, an NMR tag for high-molecular weight systems and measurements of submicromolar ligand affinities. *J Am Chem Soc* **137**, 4581–4586.

2 Loh CT, Adams LA, Graham B & Otting G (2018) Genetically encoded amino acids with *tert*-butyl and trimethylsilyl groups for site-selective studies of proteins by NMR spectroscopy. *J Biomol NMR* **71**, 287–293.

3 Liu Q, He QT, Lyu X, Yang F, Zhu ZL, Xiao P, Yang Z, Zhang F, Yang ZY, Wang XY *et al.* (2020) DeSiphering receptor core-induced and ligand-dependent conformational changes in arrestin via genetic encoded trimethylsilyl ^{1}H-NMR probe. *Nat Commun* **11**, 4857.

4 Abdelkader EH, Qianzhu H, Tan YJ, Adams LA, Huber T & Otting G (2021) Genetic encoding of N^{6}-((((trimethylsilyl)methoxy)carbonyl)-L-lysine for NMR studies of protein–protein and protein–ligand interactions. *J Am Chem Soc* **143**, 1133–1143.

5 Jackson JC, Hammill JT & Mehl RA (2007) Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterising protein structure and reactivity. *J Am Chem Soc* **129**, 1160–1166.

6 Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B *et al.* (2008) *In vivo* incorporation of unnatural amino

acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc* **130**, 9268–9281.
7 Qianzhu H, Welegedara AP, Williamson H, McGrath AE, Mahawaththa MC, Dixon NE, Otting G & Huber T (2020) Genetic encoding of *para*-pentafluorsulfanyl phenylalanine: a highly hydrophobic and strongly electronegative group for stable protein interactions. *J Am Chem Soc* **142**, 17277–17281.
8 Orton HW, Qianzhu H, Abdelkader EH, Habel EI, Tan YJ, Frkic RL, Jackson CJ, Huber T & Otting G (2021) 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* **143**, 19587–19598.
9 Qianzhu H, Abdelkader EH, Herath ID, Otting G & Huber T (2022) Site-specific incorporation of 7-fluoro-L-tryptophan into proteins by genetic encoding to monitor ligand binding by ^{19}F NMR spectroscopy. *ACS Sens* **7**, 44–49.
10 Qianzhu H, Abdelkader EH, Otting G & Huber T (2024) Genetic encoding of fluoro-L-tryptophans for site-specific detection of conformational heterogeneity in proteins by NMR spectroscopy. *J Am Chem Soc* **146**, 13641–13650.
11 Qianzhu H, Tan YJ, Abdelkader EH, Huber T & Otting G (2025) Genetic encoding of fluorinated analogues of phenylalanine for ^{19}F NMR spectroscopy: detection of conformational heterogeneity in flaviviral NS2B-NS3 proteases. *ACS Sens* **10**, 3152–3161.
12 Ekanayake KB, Mahawaththa MC, Qianzhu H, Abdelkader EH, George J, Ullrich S, Murphy RB, Fry SE, Johansen-Leete J, Payne RJ *et al.* (2023) Probing ligand binding sites on large proteins by NMR spectroscopy of genetically encoded non-canonical amino acids. *J Med Chem* **66**, 5289–5304.
13 Sigal N, Delius H, Kornberg T, Gefter ML & Alberts B (1972) A DNA-unwinding protein isolated from *Escherichia coli*: its interaction with DNA and with DNA polymerases. *Proc Natl Acad Sci USA* **69**, 3537–3541.
14 Su XC, Wang Y, Yagi H, Shishmarev D, Mason CE, Smith PJ, Vandevenne M, Dixon NE & Otting G (2014) Bound or free: interaction of the C-terminal domain of *Escherichia coli* single-stranded DNA-binding protein (SSB) with the tetrameric core of SSB. *Biochemistry* **53**, 1925–1934.
15 Jabar S, Adams L, Wang Y, Aurelio L, Graham B & Otting G (2017) Chemical tagging with *t*-butyl and trimethylsilyl groups for measuring intermolecular NOEs in a large protein-ligand complex. *Chem Eur J* **23**, 13033–13036.
16 Lohman TM & Ferrari ME (1994) *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu Rev Biochem* **63**, 527–570.
17 Shereda RD, Kozlov AG, Lohman TM, Cox MM & Keck JL (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* **43**, 289–318.
18 Shishmarev D, Wang Y, Mason CE, Su XC, Oakley AJ, Graham B, Huber T, Dixon NE & Otting G (2014) Intramolecular binding mode of the C-terminus of *Escherichia coli* single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Res* **42**, 2750–2757.
19 Raghunathan S, Kozlov AG, Lohman TM & Waksman G (2000) Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat Struct Biol* **7**, 648–652.
20 Antony E & Lohman TM (2019) Dynamics of *E. coli* single stranded DNA binding (SSB) protein-DNA complexes. *Semin Cell Dev Biol* **86**, 102–111.
21 Lohman TM & Overman LB (1985) Two binding modes in *Escherichia coli* single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration. *J Biol Chem* **260**, 3594–3603.
22 Bujalowski W & Lohman TM (1986) *Escherichia coli* single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry* **25**, 7799–7802.
23 Bujalowski W, Overman LB & Lohman T (1988) Binding mode transitions of *Escherichia coli* single strand binding protein-single-stranded DNA complexes. Cation, anion, pH, and binding density effects. *J Biol Chem* **263**, 4629–4640.
24 Kitevski-LeBlanc JL & Prosser RS (2012) Current applications of ^{19}F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* **62**, 1–33.
25 Liu JJ, Horst R, Katritch V, Stevens RC & Wüthrich K (2014) Biased signaling pathways in β2-adrenergic receptor characterized by ^{19}F-NMR. *Science* **335**, 1106–1110.
26 Mukai T, Hoshi H, Ohtake K, Takahashi M, Yamaguchi A, Hayashi A, Yokoyama S & Sakamoto K (2015) Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci Rep* **5**, 9699.
27 Guo L-T, Wang Y-S, Nakamura A & Söll D (2014) Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc Natl Acad Soc USA* **111**, 16724–16729.
28 Abdelkader EH, Qianzhu H, George J, Frkic RL, Jackson CJ, Nitsche C, Otting G & Huber T (2022) Genetic encoding of cyanopyridylalanine for in-cell protein macrocyclization by the nitrile–aminothiol click reaction. *Angew Chem Int Ed* **61**, e202114154.
29 Abdelkader EH, Qianzhu H, Huber T & Otting G (2023) Genetic encoding of 7-aza-L-tryptophan:

isoelectronic substitution of a single CH-group in a protein for a nitrogen atom for site-selective labeling. *ACS Sens* **8**, 4402–4406.
30 Kim K & Cole PA (1998) Kinetic analysis of a protein tyrosine kinase reaction transition state in the forward and reverse directions. *J Am Chem Soc* **120**, 6851–6858.
31 Hull WE & Sykes BD (1974) Fluorotyrosine alkaline phosphatase. ^{19}F nuclear magnetic resonance relaxation times and molecular motion of the individual fluorotyrosines. *Biochemistry* **13**, 3431–3437.
32 Klein-Seetharaman J, Getmanova EV, Loewen MC, Reeves PJ & Khorana HG (1999) NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: applicability of solution ^{19}F NMR. *Proc Natl Acad Sci USA* **96**, 13744–13749.
33 Luchette PA, Prosser RS & Sanders CR (2002) Oxygen as a paramagnetic probe of membrane protein structure by cysteine mutagenesis and ^{19}F NMR spectroscopy. *J Am Chem Soc* **124**, 1778–1781.
34 Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL *et al.* (2015) Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell* **161**, 1101–1111.
35 Edwards JM, Derrick JP, van der Walle CF & Golovanov AP (2018) ^{19}F NMR as a tool for monitoring individual differentially labeled proteins in complex mixtures. *Mol Pharm* **15**, 2785–2796.
36 Noren CJ, Anthony-Cahill SJ, Griffith MC & Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188.
37 Furter R (1998) Expansion of the genetic code: site-directed *p*-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci* **7**, 419–426.
38 Minnihan EC, Young DD, Schultz PG & Stubbe J (2011) Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J Am Chem Soc* **133**, 15942–15945.
39 Peeler JC & Mehl RA (2011) Site-specific incorporation of unnatural amino acids as probes for protein conformational changes. *Methods Mol Biol* **794**, 125–134.
40 Sharaf NG, Ishima R & Gronenborn AM (2016) Conformational plasticity of the NNRTI-binding pocket in HIV-1 reverse transcriptase: a fluorine nuclear magnetic resonance study. *Biochemistry* **55**, 3864–3873.
41 Chambers SE, Lau EY & Gerig JT (1994) Origins of fluorine chemical shifts in proteins. *J Am Chem Soc* **116**, 3603–3604.
42 Maxwell M, Tan YJ, Lee R, Huber T & Otting G (2023) Electrostatic contribution to ^{19}F chemical shifts in fluorotryptophans in proteins. *Biochemistry* **62**, 3255–3264.
43 Augspurger J, Pearson JG, Oldfield E, Dykstra CE, Park KD & Schwartz D (1992) Chemical-shift ranges in proteins. *J Magn Reson* **100**, 342–357.
44 Wang X, Liu D, Shen L, Li F, Li Y, Yang L, Xu T, Tao H, Yao D, Wu L *et al.* (2021) A genetically encoded F-19 NMR probe reveals the allosteric modulation mechanism of cannabinoid receptor 1. *J Am Chem Soc* **143**, 16320–16325.
45 Boeszoermenyi A, Chhabra S, Dubey A, Radeva DL, Burdzhiev NT, Chanev CD, Petrov OI, Gelev VM, Zhang M, Anklin C *et al.* (2019) Aromatic ^{19}F-^{13}C TROSY: a background-free approach to probe biomolecular structure, function, and dynamics. *Nat Methods* **16**, 333–340.
46 Boeszoermenyi A, Radeva DL, Schindler S, Valadares V, Padmanabha Das KM, Dubey A, Viennet T, Schmitt M, Kast P, Gelev VM *et al.* (2025) Leveraging relaxation-optimized ^{1}H–$^{13}C_F$ correlations in 4-^{19}F-phenylalanine as atomic beacons for probing structure and dynamics of large proteins. *Nat Chem* **17**, 835–846.
47 Huang Y, Reddy KD, Bracken C, Qiu B, Zhan W, Eliezer D & Boudker O (2023) Environmentally ultrasensitive fluorine probe to resolve protein conformational ensembles by ^{19}F NMR and cryo-EM. *J Am Chem Soc* **145**, 8583–8592.
48 Tan YJ, Abdelkader EH, Tarcoveanu E, Maleckis A, Nitsche C & Otting G (2024) (2*S*,4*S*)-5-Fluoroleucine, (2*S*,4*R*)-5-fluoroleucine, and 5,5′-difluoroleucine in *E. coli* PpiB: protein production, ^{19}F NMR, and ligand sensing enhanced by the gauche effect. *Biochemistry* **63**, 1376–1387.
49 Frere GA, Hasabnis A, Francisco CB, Suleiman M, Alimowska O, Rahmatullah R, Gould J, Su CYC, Voznyy O, Gunning PT *et al.* (2024) Next-generation tags for fluorine nuclear magnetic resonance: designing amplification of chemical shift sensitivity. *J Am Chem Soc* **146**, 3052–3064.
50 Andrews LJ & Keefer RM (1950) Cation complexes of compounds containing carbon−carbon double bonds. VI. The argentation of substituted benzenes. *J Am Chem Soc* **72**, 3113–3116.
51 Li XJ, Shan G, Liu H & Wang ZY (2009) Determination of lgK_{ow} and QSPR study on some fluorobenzene derivatives. *Chinese J Struct Chem* **28**, 1236–1241.
52 Li M, Xu G, Gong Z, Wu Q, Jiang L & Li C (2025) Simultaneous measurement of multiple fluorine labelling effect on GB1 stability by ^{19}F NMR. *Talanta* **292**, 127959.
53 Hanson GS & Coxon CR (2024) Fluorinated tags to study protein conformation and interactions using ^{19}F NMR. *Chembiochem* **25**, e202400195.

54 Yang X, Su X-C & Xuan W (2024) Genetically encoded proteinogenic and non-proteinogenic amino acids. *Chembiochem* **25**, e202400393.
55 van den Elzen PJ, Konings RN, Veltkamp E & Nijkamp HJ (1980) Transcription of bacteriocinogenic plasmid CloDF13 in vivo and in vitro: structure of the cloacin immunity operon. *J Bacteriol* **144**, 579–591.
56 Lammers C, Hahn LE & Neumann H (2014) Optimized plasmid systems for the incorporation of multiple different unnatural amino acids by evolved orthogonal ribosomes. *Chembiochem* **15**, 1800–1804.
57 Qi R & Otting G (2019) Mutant T4 DNA polymerase for easy cloning and mutagenesis. *PLoS One* **14**, e0211065.
58 Plateau P & Guéron M (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J Am Chem Soc* **104**, 7310–7311.
59 Crowley PB, Kyne C & Monteith WB (2012) Simple and inexpensive incorporation of ^{19}F-tryptophan for protein NMR spectroscopy. *Chem Commun* **48**, 10681–10683.
60 Lilkova E (2015) The PyMOL Molecular Graphics System, Version 2.0; Schrödinger, LLC.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Intact protein mass spectra of SSB mutants made with TMSK.

Fig. S2. Intact protein mass spectra of SSB P24/TMSNK.

Fig. S3. Intact protein mass spectra of SSB mutants made with fluorinated ncAAs.

Table S1. Mutations found in G1PylRS variants selected for activity with 2,6-diFTyr and 3,5-diFTyr.

Table S2. DNA and corresponding amino acid sequences of the protein used.