Functional genetic encoding of sulfotyrosine in mammalian cells

Xinyuan He1, Yan Chen2,5, Daisy Guiza Beltran3,5, Maia Kelly2,5, Bin Ma2, Justin Lawrie2, Feng Wang4, Eric Dodds2, Limei Zhang3, Jiantao Guo2 & Wei Niu1

Protein tyrosine O-sulfation (PTS) plays a crucial role in extracellular biomolecular interactions that dictate various cellular processes. It also involves in the development of many human diseases. Regardless of recent progress, our current understanding of PTS is still in its infancy. To promote and facilitate relevant studies, a generally applicable method is needed to enable efficient expression of sulfoproteins with defined sulfation sites in live mammalian cells. Here we report the engineering, in vitro biochemical characterization, structural study, and in vivo functional verification of a tyrosyl-tRNA synthetase mutant for the genetic encoding of sulfotyrosine in mammalian cells. We further apply this chemical biology tool to cell-based studies on the role of a sulfation site in the activation of chemokine receptor CXCR4 by its ligand. Our work will not only facilitate cellular studies of PTS, but also paves the way for economical production of sulfated proteins as therapeutic agents in mammalian systems.
First reported in 1954, protein tyrosine O-sulfation (PTS) entails the transfer of a sulfate group from 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to a protein tyrosine residue under the catalysis of membrane-bound tyrosylprotein sulfotransferases (TPSTs) in the trans-Golgi network. It is considered the most common type of tyrosine modification in nature and occurs exclusively on secreted and membrane-bound proteins that transit the trans-Golgi network. The introduction of negatively charged sulfate groups plays a crucial role in extracellular molecular interactions that dictate various cellular processes, including cell adhesion, leukocyte trafficking, hormone activities, and immune responses. Tyrosine-sulfated proteins (referred hereafter as sulfoproteins) are also involved in the development of various infectious diseases (e.g., AIDS and malaria), cancers, and immune-mediated diseases (e.g., allergy and rheumatoid arthritis). Hence, PTS could emerge as an important therapeutic target for the treatment of human diseases. However, in comparison to protein tyrosine phosphorylation, much of the fundamental questions of PTS remain unanswered. Indeed, the lack of understanding on the role of PTS in mammalian cell biology severely hindered the efforts to unravel potential link(s) between PTS and human diseases.

The key to a successful functional study of a PTS event is the ability to synthesize site-specifically sulfated proteins. Although sulfoproteins or sulfopeptides (fused to secreted proteins) can be synthesized in eukaryotic cells with normal or enhanced TPST activities, this method lacks the resolution that is required to investigate a specific tyrosine O-sulfation event within a protein containing multiple sulfotyrosine (sTyr) residues. As such, chemically synthesized sulfopeptides became an important tool to model the structural and functional role of PTS. Regardless of recent progress, the preparation of sTyr-containing peptides is still challenging. Furthermore, sulfopeptides may not faithfully resemble the native conformation and biological function of full-length sulfoproteins. To meet above challenges, the genetic incorporation of sTyr into proteins in response to amber nonsense codon (TAG) was reported. This method enables the synthesis of site-specifically sulfated proteins in Escherichia coli using engineered Methanocaldococcus jannaschii tRNA (MjRNA) and its cognate, sTyr-specific tyrosyl-tRNA synthetase (sTyrRS) pair. However, the reported MjRNA-sTyrRS pair is limited to E. coli host and cannot be used in mammalian cells where protein sulfation happens and plays important biological roles. While our manuscript was under preparation, Chatterjee et al. published an elegant work on the genetic incorporation of sTyr in mammalian cells.

Here we report the synthesis of site-specifically sulfated proteins in both live yeast and mammalian cells through the genetic incorporation of sTyr using an engineered tyrosyl-tRNA synthetase derived from E. coli (EcTyrRS). Unlike the recently published work by Chatterjee et al., our engineering of EcTyrRS is conducted by using yeast as the selection host. Although E. coli is a more widely adopted platform in the directed evolution of aminoaacyl-tRNA synthetases (aaRSs), yeast has the translational machinery that is more similar to that of the mammalian cells. Besides an indirect fluorescence-based assay in live cells and mass spectrometry analysis, we also conduct in vitro enzyme assays that provide direct activity profiles of evolved sTyrRSs. In order to understand the shifted substrate specificity, we obtain a crystal structure of the evolved sTyrRS in complex with sTyr, which also shed lights on future engineering of the enzyme for substrates with similar physicochemical properties. Finally, we perform a cellular and functional study of a co-translationally synthesized sulfoprotein in live mammalian cells. The study focuses on the role of PTS in chemokine signaling, which is central to chronic inflammatory conditions and is associated with the development of many human diseases. In addition to fundamental biochemical studies, this work also opens a door to the production of sulfoproteins as therapeutic agents in eukaryotes.

### Results

**Library construction and selection.** The crystal structure of E. coli tyrosyl-tRNA synthetase (EcTyrRS) in complex with L-tyrosine (PDB ID: 1X8X) was chosen in an initial protein modeling effort to identify key residues that may lead to shifted substrate specificity. We first rebuilt the ligand in the active site by adding a sulfate group to the hydroxyl group of tyrosine. The simulation revealed that four active site residues (Y37, L71, W129, and D182) in a pETTyrRS-lib vector, which also encoded an amber suppressor tRNA derived from E. coli tyrosyl-tRNA (EcTyrNA_CUA). Using the yeast homologous recombination cloning method, a library of 1 × 10^7 clones was obtained, which represented an over 99% coverage of the theoretical library diversity. The quality of the obtained library was verified by DNA sequencing.

Selection of sTyr-specific EcTyrRS variants adapted an approach that was previously established in Saccharomyces cerevisiae (Supplementary Fig. 1c). The selection host MaV203 contained chromosomal modifications that allowed the expression of a reporter enzyme, URA3, under the regulation of the transcriptional activator GAL4, which was encoded by a mutant gene containing amber codons at positions 44 and 110. Plasmid DNA of the EcTyrRS library was transformed into MaV203 together with the GAL4-encoding plasmid, pGADGAL4. In a positive selection, yeast cells were cultured in the presence of sTyr (1 mM). Effective amber suppression by EcTyrRS mutants that can charge the amber suppressor EcTyrNA_CUA with sTyr or any one of the 20 natural amino acids enabled the growth of MaV203 host on media without uracil supplementation. In a negative selection, yeast cells were cultured in the absence of sTyr. Undesirable aminoacylation by EcTyrRS mutants that can convert the exogenously added 5-fluoroorotic acid (5-FOA, 0.1%) into a cytotoxic compound and caused cell death. A total of 1 × 10^7 library cells was subjected to positive selection. Survived cells were then applied to the negative selection. Single colonies resulting from the two rounds of selection were further replicated on screening plates to verify sTyr-dependent growth in the absence of uracil and good cell fitness in the presence of 5-FOA. Four clones were confirmed to have the correct phenotype. DNA sequencing revealed two unique mutants (c1 and c2 in Fig. 1a). Positions Leu71 and Asp182 showed a complete convergence to Val and Gly, respectively, in both mutants. While Tyr37 did not change, position Trp129 either maintained the wild-type sequence or changed to Phe.

**Characterization of candidate sTyr tRNA synthetases (sTyrRSs).** We first evaluated the activity of obtained sTyrRS candidates by measuring amber suppression efficiency in yeast cells. Isolated library plasmids encoding c1 or c2 were transformed into host strain MaV203 together with a reporter plasmid, pyeGFP-N149TAG, which harbored a yeast enhanced green fluorescent protein gene (yeGFP) with an amber mutation at the N149 position. The fluorescence intensity of yeGFP was compared when cells were cultured without or with sTyr (Fig. 1b). The two sTyrRS variants led to similar levels of background fluorescence.
in the absence of sTyr and comparable yeGFP expression in the presence of 5 mM sTyr. The suppression of the amber codon in pyeGFP-N149TAG suggested that both c1 and c2 were able to aminoacylate EctRNA_{CUA} with sTyr in yeast. Furthermore, fluorescence in cells expressing c2 reached saturation at 1 mM of sTyr (Fig. 1b), which indicated better substrate recognition.

To further compare the biochemical properties of the two sTyrRS variants, we characterized them by in vitro assays. Each enzyme was expressed and purified as a fusion protein with a 6xHis tag on the C terminus. Such modification has previously been reported to have a negligible effect on the function of the wild-type EcTyrRS. The aminoacyl-tRNA synthetase activities were measured by monitoring the release of pyrophosphate in the amino acid activation step of the tRNA acylation reaction. The assays were carried out in the presence of Tyr or sTyr as the substrate (Fig. 1c). The c1 variant showed similar levels of activities on the two substrates at both 0.1 mM and 1 mM. Meanwhile, the c2 variant was more specific for sTyr. At the same concentration of 0.1 mM, the enzyme had no detectable activity on Tyr, but was active on sTyr. It also showed a fourfold higher activity on sTyr than Tyr when the substrate concentrations were increased to 1 mM. Overall, the c2 variant had better substrate specificity and maintained similar levels of activities towards sTyr as c1. It was chosen as the sTyrRS in following studies.

Evaluation of sTyrRS in mammalian cells. We examined the evolved sTyrRS for its ability to incorporate sTyr in mammalian cells using a two-plasmid method. The sTyrRS-encoding gene was inserted behind a non-regulated CMV promoter on plasmid pEGFP37, which has one copy of the amber suppressor tRNA under the control of a human U6 promoter. The reporter plasmid, pEGFP, contains an EGFP gene with an amber mutation at the Y40 position and DNA sequence that encoded a C-terminus 6xHis tag. The selective incorporation of sTyr into proteins was tested in 293 T cells that were transiently transfected with plasmids pEGFP37 and pEGFP. Following 24 h of incubation in media without and with sTyr (1 mM), the amber suppression level was studied by visualizing the expression of fluorescent protein under confocal microscope (Fig. 2a). Significant amount of EGFP was only expressed in cells supplemented with sTyr, while minimal background was observed without sTyr (Fig. 2a).

In addition to cell imaging, flow cytometry was used to quantify the amber suppression efficiency based on the fluorescence intensity (Supplementary Fig. 2). As shown in Fig. 2b, when the suppression efficiency of the wild-type EcTyrRS was assigned as 100%, the efficiency of sTyrRS is about 68%, which is significantly higher (P < 0.01) than a well-established EcTyrRS mutant, AzFRS (for the incorporation of p-azido-L-phenylalanine; AzF). On the other hand, the evolved sTyrRS did show a higher background signal than AzFRS when noncanonical amino acid (ncAA) was not available (Fig. 2b). We also conducted protein expression of an EGFP mutant that contains two amber codons in its encoding gene at positions Y40 and N150 (on plasmid pEGFP-double). As shown in Supplementary Fig. 3, EGFP fluorescence was only detected when 1 mM sTyr was included in the cell culture media. The expression level of EGFP-double (Supplementary Fig. 3) was apparently lower than that of EGFP containing one amber codon (Fig. 2a).

Finally, the fidelity of the evolved sTyrRS was quantified by mass spectrometry. Since the sulfoester bond in sTyr is labile and can readily decompose to yield Tyr under mass spectrometry conditions, it would be impossible to directly analyze the ratio of sTyr versus mis-incorporated Tyr in the target protein. To solve this problem, we synthesized and used di-deuterium labeled sTyr (sTyr-D2, Supplementary Fig. 4a) in the incorporation study. To reduce the cost of this experiment, we used a mixture of sTyr-D2 and unlabeled sTyr instead of pure sTyr-D2. The mixture contained sTyr-D2 at 24.2 ± 0.3% (mol%) by mass spectrometry quantification. To conduct the analysis, the EGFP-Tyr40TAG was expressed in 293T cells in the presence of the sTyr-D2/sTyr mixture at 1 mM. Since the cell culture media (DMEM with 10% FBS) contains at least 0.4 mM Tyr (a component of DMEM), we should see a significant dilution of the deuterium label at position 40 of EGFP if the evolved sTyrRS incorporates Tyr at a substantial level. The mutant protein was purified by Ni affinity resin followed by SDS-PAGE (Supplementary Fig. 4b), then subjected to trypsin digest. The resulting peptides were analyzed by a Synapt G2S HDMS quadrupole time of flight (Q-TOF) hybrid mass spectrometer. The peptide fragment that contained the position of incorporation showed the percentage of sTyr-D2 at position 40 as 24.3 ± 1.3%, which is in close agreement with the percentage of di-deuterium labeled sTyr supplemented to the cell culture media (Supplementary Fig. 5). The data excluded misincorporation of unlabeled Tyr from the cell culture media at position 40 of EGFP in the presence of sTyr. Otherwise, it would lead to the dilution effect and lower the percentage of deuterium labeled peptide. While some background incorporation of Tyr was observed in the absence of sTyr (Fig. 2a), the above studies demonstrate that the evolved sTyrRS can aminoacylate amber suppressor tRNA with high efficiency and fidelity in the presence of sTyr in mammalian cells.
Structural insights into the evolved sTyrRS. To understand key changes that shift the substrate specificity of the evolved sTyrRS, we conducted structural study of its amino acid activation domain (a.a. 1-322). The domain was overexpressed and purified as a fusion protein with an N-terminal 6xHis-SUMO tag, which was removed by treatment with the SUMO protease prior to crystallization. We obtained and refined the crystal structure of the evolved sTyrRS in complex with sTyr at 1.78 Å resolution. The protein formed a homodimer, the same as the wild-type EcTyrRS (Supplementary Fig. 6a)29. In comparison to the structure of the wild-type EcTyrRS in complex with L-Tyr (PDB ID: 1X8X), the introduction of the L71V and D182G mutations in the evolved sTyrRS led to an expanded active site to accommodate the larger side chain of sTyr (Fig. 3a, b). In addition, the D182G mutation eliminates the potential repulsion between the negatively charged carboxylate group of Asp and the sulfate group of sTyr. This mutation also contributed to the lowered affinity of the evolved sTyrRS towards Tyr, due to the abolishment of a stabilizing hydrogen bond (Fig. 3d). A third mutation, W129F, contributes to the substantially lower background activity of the c2 variant towards Tyr. This is likely due to the smaller size of the Phe side chain, which further contributes to the enlarged active site (Fig. 3a). A close inspection of ligand binding revealed that the phenol ring of the sTyr tilted around 26° away from that of the Tyr (Supplementary Fig. 6f). This movement enabled the formation of a hydrogen bond between the sulfate group and the hydroxyl group of Tyr37 residue (Fig. 3c). Since this Tyr is also critical for the binding of Tyr substrate in the active site of the wild-type EcTyrRS (Fig. 3d), its retention likely played a major role in the background activity of the evolved sTyrRS towards Tyr when sTyr was not available. Another key interaction that anchors the sulfate group is a hydrogen bond with the side chain of Asn126 (Fig. 3c), which does not participate in substrate binding in the wild-type EcTyrRS. Finally, the three mutations in the evolved sTyrRS did not interrupt the three hydrogen bonds that are essential for recognizing the α-amino group of the substrate (Supplementary Fig. 6d and 6e).

Functional incorporation of sTyr in CXCR4 protein by sTyrRS. To demonstrate that sTyr can be functionally and selectively incorporated into naturally sulfated positions of target proteins in live mammalian cells, we conducted studies of chemokine receptor CXCR4. The activation of CXCR4 by its cognate chemokine ligand, stromal cell-derived factor 1 (SDF-1; also known as CXCL12), triggers multiple downstream signal transduction pathways that regulate intracellular calcium flux, chemotaxis, transcription, and cell survival. It is known that sulfation of tyrosine residues (Y7, Y12, and Y21, Supplementary Fig. 7a) near the N-terminus of CXCR4 is required for CXCR4 activation.
for optimal binding with SDF-1. According to previously reports, the relative importance of the three sulfation sites of CXCR4 for its ligand binding and activation is Tyr21 > Tyr12 > Tyr \textsuperscript{740} \textsuperscript{41}. Therefore, we decided to focus on the Tyr21 site of CXCR4 in this work.

The incorporation of sTyr at the 21 position of CXCR4 was examined using plasmids psTyrRS and pCXCR4(21TAG)-EGFP, which encoded a CXCR4 mutant as a fusion protein with EGFP on its C terminus. The suppression of the amber mutation in CXCR4 led to the expression of EGFP in 293 T cells of which the fluorescence was monitored by flow cytometry. In the presence of sTyr, a twofold increase in fluorescence was observed, which was 43.5% of the suppression efficiency when the wild-type EcTyrRS was used (Supplementary Fig. 7b). To further confirm that sTyr was specifically incorporated at position 21, we also conducted western blot analyses. As a membrane protein, CXCR4 showed anomalous migration in SDS-PAGE \textsuperscript{42}. To circumvent this issue, western blot analyses. As a membrane protein, CXCR4 showed (Supplementary Fig. 7c). Protein with His-tag was partially we fused the DNA sequence that encodes the antibodies (Fig. 4b). The sample from cells cultured with sTyr showed positive signals with both antibodies. In contrast, the sample from cells cultured without sTyr was only positive with the anti-histidine antibody (Fig. 4b). As controls, purified pCXCR4(N38-FF 21TAG)-EGFP plasmid also included Y7F, Y12F and an amber mutation at the Tyr21 position. The Y7F and Y12F mutations eliminate any possibility of TPST-catalyzed sulfation at these two sites and makes the position 21 the only site that contains sTyr residue through the sTyrRS-mediated incorporation, which simplifies the interpretation of western blot experiments with the anti-sTyr antibody.

Probing lysates of 293T cells transfected with plasmids pCXCR4(N38-DD 21TAG)-EGFP and psTyrRS with anti-histidine tag antibody showed the higher expression level of a protein corresponding to the size of CXCR4(N38 FF 21sTyr)-EGFP when sTyr (1 mM) was provided to cells (Fig. 4a, Supplementary Fig. 7d) and further subject to western blot using both anti-histidine tag and anti-sTyr antibodies (Fig. 4b). The sample from cells cultured with sTyr showed positive signals with both antibodies. In contrast, the sample from cells cultured without sTyr was only positive with the anti-histidine antibody (Fig. 4b). As controls, purified CXCR4(N38)-EGFP and CXCR4(N38-DD)-EGFP, which had Tyr to Phe mutations at all three positions, were also only detected by anti-histidine tag antibody. Since the CXCR4(N38)-EGFP variants were cytosolic, they did not transit through trans-Golgi network where PTSTs reside and were not sulfated by the endogenous protein sulfation machinery. Results of the above western blot analyses confirmed that sTyr was successfully incorporated in response to the amber codon at position 21.

Next, we examined the function of translationally incorporated sTyr at position 21 of CXCR4 using the calcium mobilization assay. To eliminate the background signal caused by the expression of endogenous CXCR4 in 293T cell line, we generated a CXCR4-knockout cell line, 293T-ΔCXCR4 (C4), using CRISPR/Cas method (see details in Supplemental Information) \textsuperscript{43}. An approximate 90% reduction of signal in calcium mobilization assay was observed when C4 was treated with SDF-1 (Fig. 4c). The signal in C4 was restored when the wild-type CXCR4 was expressed from plasmid pCXCR4 and reduced back to near basal level when a CXCR4-FFF mutant, which had Y7F, Y12F, and Y21F mutations, was expressed (Fig. 4c). The tyrosine to phenylalanine mutation has been widely used to remove sulfation sites with minimal structural perturbation to target proteins,\textsuperscript{9,11,44,45}. To specifically probe the functional incorporation of sTyr at position 21 in CXCR4, calcium mobilization in C4 cells expressing CXCR4(FF TAG) (Y7F, Y12F, and Y21TAG) was studied. As shown in Fig. 4c, SDF-1-triggered calcium mobilization was observed when CXCR4(FF TAG) was co-expressed with the evolved sTyrRS in the presence of 1 mM sTyr. On the other hand, over twofold lower and near basal level calcium mobilization was detected when CXCR4(FF TAG) was expressed in the absence of the evolved sTyrRS (p < 0.05). The above results validated a previous finding,\textsuperscript{40,41} that sulfation at Tyr21 position of CXCR4 is important for the activation of this receptor by its chemokine ligand. Overall, our work demonstrated that genetically incorporated sTyr can reconstitute the function of CXCR4 and is suitable for cellular studies.

**Discussion**

By using yeast as the selection host, we have identified EcTyrRS mutants that were able to genetically encode sTyr in both yeast and mammalian cells. To overcome the problem of low transformation efficiency with circular DNA, linear DNA fragments were transformed into yeast cells and assembled into EcTyrRS library through the host’s homologous recombination machinery. We could regularly obtain libraries of 10\textsuperscript{7} diversity, which is comparable to the capacity of the commonly used E. coli platform. This would facilitate current efforts on engineering...
aaRSs for genetic incorporation of ncAAs into proteins in eukaryotic cells. The evolved sTyrRS displayed good efficiency in comparison to its parental wild-type EcTyrRS and to a well-established AzFRS. While a side-by-side evaluation has not been conducted with evolved sTyrRSs from this and Chatterjee’s work, comparable activities are expected since they have similar mutations, including Leu71Val and Asp182Gly. The Chatterjee’s sTyrRS does not have the Trp129Phe mutation as the sTyrRS in this report, but does contain a unique Leu186Met mutation. Residue Leu186 sits on the same α helix as residue Asp182, but is further away from the ligand in the wild-type EcTyrRS (Supplementary Fig. 8a). In the expanded substrate pocket of our mutant sTyrRS, Leu186 resides on the active site surface and mutation at this site, similar to our Trp129Phe mutation, may fine tune the sTyr binding pocket (Supplementary Fig. 8b). Incorporation of the Leu186Met or other mutations at this position of our sTyrRS mutant may lead to further improvement of properties. The evolved sTyrRS displayed certain level of background. However, mis-incorporation of Tyr was not observed in the presence of sTyr, which suggests that the evolved sTyrRS is more active on sTyr than Tyr. This notion was confirmed by the results of in vitro enzyme assays. Our structural study showed that Tyr37 could form a hydrogen bond with either sTyr or Tyr. It therefore can potentially contribute to the background activity of the evolved sTyrRS towards Tyr. Further mutagenesis that weakens hydrogen bonding between Tyr37 and the undesirable Tyr substrate can likely lead to sTyrRS variants with higher specificity. We are currently conducting molecular dynamics simulations using the obtained sTyrRS structure to identify potential sites for modification. Our findings will be reported in due course.

In a proof of concept study using the CXCR4/SDF-1 axis as the model system, we demonstrated that the genetically incorporated sTyr could be readily used to study the role of sTyr in the activation of chemokine receptors in live mammalian cells. While a single sulfation site (sTyr21) was the focus of this work to demonstrate the utility of genetic encoding of sTyr in live-cell studies, it is possible to investigate multiple sulfation sites simultaneously in the future. However, caution is recommended in such studies since the efficiency of genetic incorporation of sTyr is likely decreased due to the need to suppress multiple amber codons within the same transcript. Since many G protein-coupled receptors, including chemokine receptors, are sulfated, it is crucial to understand whether and how sulfation events affect receptor-ligand interactions. The system established in this work is a useful tool to facilitate the study of protein sulfation-associated eukaryotic signaling processes and the development of diagnostic and therapeutic agents targeting protein sulfation. Furthermore, the genetic incorporation of sTyr is expected to enable facile productions of therapeutic proteins containing tyrosine sulfation, which has been reported as a key post-translational modification to boost the efficacy.

Methods
Reagents and media. All commercial chemicals are of reagent grade or higher. All solutions were prepared in deionized water that was further treated by Barnstead Nanopure® ultrapure water purification system (Thermo Fisher Scientific). Preparation of LB medium and YPD medium followed the reported recipe. Yeast strain MaV203 (MATa; leu2-3,112; trp1-109; his3Δ200; ade2-101; cyh2Δ; gln3Δ1; gln5Δ1; gln80Δ1; GAL1::lacZ; HIS3::GAL1::lacZ; HIS3::GAL1::lacZ) and ΔECs204 (MATa; leu2-3,112; trp1-109; his3Δ200; ade2-101; cyh2Δ; gln3Δ1; gln5Δ1; gln80Δ1; GAL1::lacZ; HIS3::GAL1::lacZ; HIS3::GAL1::lacZ) (Thermo Fisher Scientific) were used for yeast library construction and selection. Standard protocols were followed for the purification and analysis of plasmid DNA. PCR amplifications were carried out using KOD Hot Start DNA polymerase (MilliporeSigma) by following the manufacturer’s protocol. Other molecular cloning reagents were purchased from New England Biolabs. Primer synthesis and DNA sequencing services were performed by Eurofins MWG Operon. Primers used in this study are listed in Supplementary Table 1. Plasmids used and constructed in this study are listed in Supplementary Table 2. All plasmids are available from corresponding authors upon reasonable request.

Plasmid pET30b-sTyrRSs. These 6.7-kb plasmids were pET30b-derived plasmids that encode candidate sTyrRSs with a C-terminus 6xHis tag. DNA fragments encoding candidate sTyrRSs were amplified without the stop codon using isolated library plasmids as templates. PCR products were assembled into pET30b (Novagen) vector that was linearized by digestion with Ndel and Xhol.

Plasmid pT30b-sTyrRSs. These 6.7-kb plasmids were pET30b-derived plasmids that contain one copy of the amber suppressor R. sputilis tyrA-tRNA behind a human U6 promoter and an indicated tRNA synthetase-encoding gene behind a CMV promoter. The construction of the plasmid was completed by replacing an AzFRS-encoding gene in pCDNA3.1-AzFRS with the desired DNA sequence. Fragments encoding the wild-type EcTyrRS or the evolved sTyrRSs were amplified from E. coli genomic DNA or isolated library plasmid. Purified PCR product was assembled into a 6.2-kb DNA fragment from the digestion of pCDNA3.1-AzFRS by HindII and AplI.

Mammalian cell experiments. Mammalian cells were maintained in DMEM (Gibco®) media supplemented with 10% FBS (v/v) at 37 °C in a humidified atmosphere of 5% CO2 (v/v). Transfection was conducted at 70–80% cell confluence using Lipofectamine 2000 (Life Technologies) according to the manufacturer’s protocol. For purification of protein with 6xHis tag, cells from a T-25 or T-75 flask were detached and lysed in RIPA lysis and extraction buffer (Thermo Scientific) with protease inhibitor cocktail (EDTA-free, Thermo Scientific). Following centrifugation at 21,000 x g for 30 min at 4 °C, the supernatant was collected and dialyzed into purification buffer, which contained potassium phosphate (25 mM, pH 7.5), NaCl (200 mM), and BSA (10 mM). The tagged protein was purified using Ni Sepharose 6 Fast Flow resin (GE Healthcare) by following the manufacturer’s protocol.

Biochemical analysis. Protein concentrations were determined by using the Quick Start™ Bradford protein assay kit (Bio-Rad Laboratories). Western blot experiments started with sample preparation by SDS-PAGE. Proteins were then transferred onto a nitrocellulose membrane using mini trans-blot cell (Bio-Rad Laboratories) at 100 V for 1 h. The membranes were blocked overnight at 4 °C, then incubated with primary and secondary antibodies for 1 h each: mouse anti-6xHis tag monoclonal antibody (1:3000, MCA1396, Bio-Rad Laboratories), mouse anti-sulfotyrosine antibody clone sulfo-1C-A2 (1:1000, 05-1100, MilliporeSigma), and HRP conjugated goat anti-mouse IgG (1:1000, 1705616, BioRad Laboratories). Blots were developed using Opti-4CN detection kit (BioRad Laboratories) and scanned by using Gel Doc XR+ software (Bio-Rad Laboratories). A TX plate reader was used in absorbance, fluorescence, and kinetics measurements.

Strains and plasmids. E. coli GeneHogs® (Thermo Fisher Scientific) was used for routine cloning and plasmid propagation. Plasmid construction was performed using Gibson Assembly method. E. coli BL21(DE3) (New England Biolabs) was used for protein expression and purification. S. cerevisiae strain MaV203 (MATa; leu2-3,112; trp1-109; his3Δ200; ade2-101; cyh2Δ; gln3Δ1; gln5Δ1; gln80Δ1; GAL1::lacZ; HIS3::GAL1::lacZ; HIS3::GAL1::lacZ; HIS3::GAL1::lacZ) (Thermo Fisher Scientific) was used for yeast library construction and selection. Standard protocols were followed for the purification and analysis of plasmid DNA. PCR amplifications were carried out using KOD Hot Start DNA polymerase (MilliporeSigma) by following the manufacturer’s protocol. Other molecular cloning reagents were purchased from New England Biolabs. Primer synthesis and DNA sequencing services were performed by Eurofins MWG Operon. Primers used in this study are listed in Supplementary Table 1. Plasmids used and constructed in this study are listed in Supplementary Table 2. All plasmids are available from corresponding authors upon reasonable request.
Plasmid pEGFP-double. The 4.3 kb plasmid is derived from pEGFP by replacing a 0.8 kb DNA fragment, which has flanking XbaI and Apal sites, with a PCR product that contains the N15TAG mutation.

Plasmid pCXCR4 and derivatives. The 7.2 kb pCXCR4 plasmid is derived from pcDNA3.1 by replacing the AzF5R-encoding gene in pcDNA3.1-AzF5R with human CXCR4 variant 2. DNA fragment encoding human CXCR4 was amplified from pcE4P4-FLAG-CXCR452 (Addgene 98947). Purified PCR product was assembled into a 6.2 kb DNA fragment from the digestion of pcDNA3.1-AzF5R by HindIII and Apal. Mutations were introduced into the CXCR4 gene of pCXCR4 plasmid by mutagenic PCR.

Plasmid pCXCR4-EGFP and derivatives. The 8.0 kb pCXCR4-EGFP plasmid is derived from pcDNA3.1 by replacing the AzF5R-encoding gene in pcDNA3.1-AzF5R with a fusion gene between the human CXCR4 variant 2 and EGFP. DNA fragments encoding human CXCR4 without the stop codon and mammalian EGFP were amplified from plasmid pCXCR4 and pEGFP, respectively. Purified PCR products were assembled into a 6.2 kb DNA fragment from the digestion of pcDNA3.1-AzF5R by HindIII and Apal. Mutations were introduced into the CXCR4 gene of pCXCR4-EGFP plasmid by mutagenic PCR.

Plasmid pCXCR4(N38)-EGFP and derivatives. The 7.0 kb pCXCR4(N38)-EGFP plasmid was constructed in a similar way to pCXCR4-EGFP, except the PCR product of DNA that encodes the first 38 amino acid of CXCR4 (CXCR4(N38)) was used in the Gibson assembly reaction. Obtained plasmid expressed a fusion protein CXCR4(N38)-EGFP. Mutations were introduced into the CXCR4(N38) sequence of pCXCR4(N38)-EGFP plasmid by mutagenic PCR.

Plasmid pET-6xHis-SUMO-sTyrRS. Plasmid pET-6xHis-SUMO was constructed into a 6.2 kb DNA fragment from the digestion of pcDNA3.1-AzF5R by HindIII and Apal. Purified PCR product was assembled into the pET-6His-SUMO vector that was linearized by BamHI and XhoI.

Library construction and selection. To construct the EcTyrRS library, four sites (Y37, L71, W129, and D182) were targeted for randomization using the NNK codon (K or T or G) by PCR of plasmid pET-hsTyrRS. Each of the four sites was amplified as a separate PCR product. The four PCR products were assembled into a 6.2 kb DNA fragment from the digestion of pcDNA3.1-AzF5R by HindIII and Apal. Mutations were introduced into the CXCR4 gene of pET-6xHis-SUMO vector that was linearized by BamHI and XhoI.

Confocal microscopy. 293T cells (1 × 10⁴) seeded in a single well of a 24-well plate were grown for 24 h, then transfected with plasmids pSTyrRS (0.8 µg) and pEGFP double in 0.5 mL of Lipofectamine. Transfected cells were cultured for an additional 24 h in 0.5 mL of media without or with sTyr (1 mM). Cells were then washed with 0.5 mL of warm DMEM base medium, fixed with 4% paraformaldehyde (w/v) for 15 min. Following removal of the fixation reagent by washing with 3 × 0.5 mL DPBS, cells were visualized by an Inverted (Olympus IX 81) confocal microscope.

Flow cytometry analysis. For the quantification of amber suppression efficiency by sTyrRS, 293T cells (1 × 10⁴) seeded in a single well of a 24-well plate were grown for 24 h, then transfected with one plasmid that encodes a rRNA synthetase (0.8 µg) of psTyrRS, AZFRS, or EcTyrRS and pEGFP (0.8 µg) in 2 µL of Lipofectamine. Transfected cells were cultured for an additional 24 h in 0.5 mL of media without or with 1 mM sTyr or AzF. Cells were detached with 0.3 mL of Trypsin/ EDTA (0.05%), washed once with 1 mL DPBS, and collected by centrifugation at 300 × g for 5 min. Cells collected in 0.5 mL 4% paraformaldehyde (w/v) and incubated for 20 min at room temperature. Following removal of the fixation reagent by centrifugation, cells were resuspended in 0.5 mL DPBS and kept on ice until analysis. Fluorescence of cells was measured using a Beckman Coulter CytoFLEX flow cytometer. A total of 30,000 cells were analyzed for each sample. Data were analyzed using FlowJo. Reported data are the average measurement of three samples with standard deviations. Flow cytometry analysis of sTyr incorporation into CXCR4(212TAG)-EGFP followed the similar method.

In vitro rRNA synthetase assay. For protein purification, plasmid pET30b-sTyrRS was transformed into protein expression host E. coli BL21 (DE3). A 1 mL overnight seed culture was introduced into 100 mL of fresh LB medium containing kanamycin. When cell growth reached exponential phase (OD₆₀₀ = 0.4–0.6), protein expression was induced with IPTG (0.5 mM). Cells were cultured continuously at room temperature for an additional 12 h prior to harvesting. Collected cells were lysed by sonication and rRNA synthetase with C-terminal 6×His tag was purified on Ni resin by following the manufacturer’s protocol, except for replacing phosphate buffer with Tris–HCl buffer (25 mM, pH 7.5). Protein concentration was measured using Bradford assay and Bradford reagent was used as a standard. Protein purity was measured using SDS-PAGE.

For rRNA substrate preparation, the E. coli tyrosyl-rRNA substrate with CUA anticodon loop (EUGNA,UA) was prepared by in vitro transcription using the MEGAscript® Kit (Life Technologies). A T7 promoter sequence (5′- TAATACGACTCACTATAGGG-3′) was added to the 5′ end of the forward primer. Primers T7-Tyr-F and T7-Tyr-R were used to amplify the EUGNA,UA from pET-TyrRS-Lib template. A 2 µL reaction was set up with 0.2 µg PCR product, 8 µL NTPs mixture (18.75 mM each of ATP, CTP, GTP, and UTP), 2 µL enzyme mix, and 2 µL reaction buffer. The reaction was incubated at 37°C for 12 h. The in vitro transcription was stopped by the addition of 150 µL ammonium acetate, 10 µL glycogen, and 10 µL EDTA (0.05%, Thermo Fisher Scientific) and mixed on ice for 5 min. Collected aqueous phase was precipitated with an equal amount of isopropanol. Obtained rRNA pellet was dissolved in 50 µL nuclease-free water and the concentration was determined by A₂₆₀. The rRNA was folded prior to aminoacylation reactions by heating at 70°C for 10 min, followed by the addition of MgCl₂ to 10 mM, then slow cooling by incubation at room temperature for 5 min.

For rRNA synthetase assay, standard in vitro aminoacylation assays were performed using EnzChek® Pyrophosphate Assay Kit (Thermo Fisher Scientific) to continuously monitor the release of PPi. The assay mixture contained ATP (0.1 mM), rRNA (10 µM), rRNA synthetase (5 µM), tyrosine or sulfotyrosine at specified concentrations, 2-amino-6-mercapto-7-methylpyrline ribonucleoside (MESG, 0.2 mM), purine nucleoside phosphorylase (1 U/mL), and inorganic pyrophosphatase (0.03 U/mL) in a reaction buffer of HEPES (50 mM, pH 7.5) with MgCl₂ (10 mM), KC (50 mM), and dithiothreitol (1 mM). Assays were carried out in 384-well format at room temperature by monitoring the absorbance change at 360 nm. Each reaction was initiated by the addition of an amino acid substrate. Control reactions without a substrate were performed. The standard curve of A₂₆₀ vs. PPi concentration was constructed and used for the calculation of rRNA synthetase activities.

Mass spectrometry analysis of sTyr incorporation in mammalian cells. For protein purification, a T7-5 flask was seeded with 3.0 × 10⁹ 293T cells. Following 24 h of growth, cells were transfected with pET30b (19.2 µg) and pEGFP (19.2 µg) in 48 µL of Lipofectamine. Cells were cultured for an additional 24 h in media containing 1 mM of sTyr or sTyr mixture. Protein EGFP-Tyr40sTyr was purified from collected cells using Ni resin and further resolved from impurity by SDS-PAGE (Supplementary Fig. 4b). The band of EGFP-Tyr40sTyr was excised, washed, and treated with 37°C Tryptic peptides were extracted from the gel pieces, dried down, and re-dissolved in 25 µL of aqueous solution of acetonitrile (2.5%) and formic acid (0.1%).
For mass spectrometry analysis, all experiments were conducted using a Synapt G2S HDMS Q-TOF hybrid mass spectrometer (Waters, Manchester, UK) equipped with a Nanospray VialPro Velocity ESI ion source (Waters, Manchester UK) and an Acquity UPLC M-Class Peptide BEH C18 Column. Separation of peptides was performed using an acetonitrile gradient (0-85% ACN over 35 minutes) in the presence of 0.1% formic acid with the flow rate of 6 μL/min. The spectral acquisition was conducted in MassLynx 4.1 (Waters, Manchester, UK) with data handling performed using DataAnalysis 6.3 (Waters, Manchester UK). Deuterated percentages were calculated using peak area integrations of the M0 and [(M0 + 2) + M1] peaks and predicted isotope abundances of the target molecules.

\[
\text{Deuterated Percentage} = \frac{M_0 + M_1}{M_0} \times 100
\]

Crystalization of sTyrRS. For protein purification, plasmid PET-6xHis-SUMO-sTyrRS was transformed into plasmid expression host E. coli BL21(DE3). Cell culturing, induction of protein expression, and purification of the tagged protein followed the same protocol as the purification of full-length sTyrRS. The N-terminal 6xHis-SUMO tag was removed by treatment with Ulp1. Following protease digest, the mixture was passed through Ni resin and the flow-through fraction, which contained unfractionated sTyrRS, was collected. Obtained protein was dialyzed into the desalting buffer that contained NaCl (50 mM) and 2-mercaptoethanol (5 mM) in Tris-HCl buffer (20 mM, pH 7.5) using 10DG desalting columns (Bio-Rad Laboratories). Protein was further purified by gel filtration using a Superdex 200 column (GE Healthcare Life Sciences) at a flow rate of 2 mL/min. Collected fractions were concentrated to 10 mg/mL, flash frozen in liquid nitrogen, then stored at −80 °C until crystalization trials.

The sTyrRS, s2L-sulfotyrosine complex was crystalized by sitting drop vapor diffusion at 15 °C. Sulfotyrosine was added to the purified sTyrRS protein (a.a. 1–322, 9 mg/mL) to a final concentration of 2 mM. Obtained protein solution was then mixed at a 2:1 ratio (v/v) with the reservoir solution containing ammonium sulfate (2.2 M) and PEG 400 (2.5 M) in 2 mM Tris-HCl buffer (pH 7.5) using 10DG desalting columns (Bio-Rad Laboratories). Protein was further purified by gel filtration using a Superdex 200 column (GE Healthcare Life Sciences) at a flow rate of 2 mL/min. Collected fractions were concentrated to 10 mg/mL, flash frozen in liquid nitrogen, then stored at −80 °C until crystallization trials.

The X-Ray diffraction data were collected at Beamline 12-2 of the Stanford Synchrotron Radiation Lightsource (SSRL), with a 6 Pixel Array Detector. A set of 1440 diffraction images were collected at 12660.4 eV with an oscillation angle of 0.25°. Diffraction data were integrated with the XDS program and subsequently processed using the CCP4 suite65. The crystal structure of Escherichia coli sTyrRS (PDB ID: 2YXN) was used as the initial model for phase determination56. Further structural refinement was carried out using PHENIX, REFMAC5 and COOT59. The structure was deposited in Protein Data Bank (PDB ID: 6WN2).

Generation of CXCR4-knockout cell line. A CXCR4-knockout cell line was generated from 29T using CRISPR/Cas9 method43. The sequence of the guide RNA was −ACTTCAGATAACTACACCG-3′. The sequence of the CRISPR/Cas9 guide RNA was −ACTTCAGATAACTACACCG-3′. A CXCR4-knockout cell line was amplified by PCR. The PCR product was cloned into pCR2.1 (Invitrogen). Cell cultures were seeded in a single well (1 × 10^5 cells) of a poly-lysine coated, clear bottom, black wall 96-well plate (Greiner Bio-One). Following incubation for 24 h at 37 °C, the culture media was replaced with 10 μL DMEM containing 4 μM Fluo-4 AM (Thermo Fisher Scientific) and 0.04% pluronic acid. Following a hour incubation at room temperature protected from light, media was replaced with fresh DMEM and cells were incubated at room temperature protected from light for another 45 min. Cells were washed with 100 μL DPBS once, then immersed in 100 μL fresh DPBS and ready for assay. Calcium mobilization was measured using excitation at 485 nm and emission at 528 nm. Fluorescence of cells was monitored for 2 min prior to the addition of SDF-1 (20 nM) and for 4 min after the stimulation. The calcium response was calculated as the percentage of fluorescence change over baseline. Data are expressed as the average of three measurements with standard deviation.

Statistics and reproducibility. Standard deviations and Student’s t-test were performed using Microsoft Excel 365 software. Similar results were obtained from three independent experiments.

Data availability
All data generated or analyzed during this study are included in this article and its Supplementary Information. All relevant data are available from the authors upon reasonable request. Source data are provided with this paper. Crystal structure of sTyrRS in complex with sTyr is deposited in Protein Data Bank (PDB ID: 6WN2). Crystal structures of E. coli tyrosyl-sRNA synthetase (sTyrRS) in complex with L-sulfotyrosine (PDB ID: 1X85) and E. coli TyrRS (PDB ID: 2YXN) are accessed from Protein Data Bank. Source data are provided with this paper.

Received: 24 April 2020; Accepted: 27 August 2020; Published online: 24 September 2020

References
1. Bettelheim, F. R. Tyrosine-O-sulfate in a peptide from fibrinogen. J. Am. Chem. Soc. 76, 2838–2839 (1954).
2. Baetzieke, P. A. & Huttner, W. B. Tyrosine sulfation of yolk proteins 1, 2, and 3 in Drosophila melanogaster. J. Biol. Chem. 260, 6434–6439 (1985).
3. Huttner, W. B. Tyrosine sulfation and the secretory pathway. Annu. Rev. Physiol. 50, 363–376 (1988).
4. Stone, M. J., Chuang, S., Hou, X., Shoham, M. & Zhu, J. Z. Tyrosine sulfation: an increasingly recognized post-translational modification of secreted proteins. N. Biotechnol. 25, 299–317 (2009).
5. Seibert, C. & Sakmar, T. P. Toward a framework for sulfoproteomics: synthesis and characterization of sulfotyrosine-containing peptides. Biopolymers 90, 459–477 (2008).
6. Hille, A., Braulke, T., van Figura, K. & Huttner, W. B. Occurrence of tyrosine sulfate in proteins - a balance sheet. 1. Secretory and lysosomal proteins. Eur. J. Biochem. 188, 577–586 (1990).
7. Hille, A. & Huttner, W. B. Occurrence of tyrosine sulfate in proteins - a balance sheet. 2. Membrane proteins. Eur. J. Biochem. 188, 587–596 (1990).
8. Woods, A. S., Wang, H.-Y. & Jackson, S. N. Sulfation, the up-and-coming post-translational modification: its role and mechanism in protein-protein interaction. J. Proteome Res. 6, 1176–1182 (2007).
9. Farzan, M. et al. Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. Cell 96, 667–676 (1999).
10. Cormier, E. G. et al. Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120. Proc. Natl. Acad. Sci. USA 97, 5762–5767 (2000).
11. Choe, H. et al. Sulphated tyrosines mediate association of chemokines and Plasmodium vivax Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). Mol. Microbiol. 55, 1413–1422 (2005).
12. Hsu, W., Rosenquist, G. L., Ansari, A. A. & Gershwin, M. E. Autoimmunity and tyrosine sulfation. Autoimmun. Rev. 4, 429–435 (2005).
13. Ballou, F. R. et al. The chemokine receptor CXCR4 is ligated by CD4+ T cells and analyzed by DNA gel electrophoresis. J. Exp. Med. 199, 1423–1434 (1999).
14. Bryant, V. L. & Slade, C. A. Chemokines, their receptors and human disease: an increasingly recognized post-translational modification of secreted proteins. N. Biotechnol. 25, 299–317 (2009).
15. Hans, D. et al. Mapping binding residues in the tyrosyl-sRNA synthetase (sTyrRS) in complex with L-sulfotyrosine (PDB ID: 1X85) and E. coli TyrRS (PDB ID: 2YXN) are accessed from Protein Data Bank. Source data are provided with this paper.
20. Choe, H. et al. Tyrosine sulfation of human antibodies contributes to recognition of the CC55 binding region of HIV-1 gp120. Cell 114, 161–170 (2003).

21. Dorfman, T., Moore, M. J., Guth, A. C., Choe, H. & Farzan, M. A tyrosine-sulfated peptide derived from the heavy-chain CDR3 region of an HIV-1-neutralizing antibody binds gp120 and inhibits HIV-1 infection. J. Biol. Chem. 281, 28529–28535 (2006).

22. Yagami, T., Shiwa, S., Futaki, S. & Kitagawa, K. Evaluation of the final deprotection system for the solid-phase synthesis of Tyr(SO3H)-containing peptides with 9-fluorenylethoxycarbonyl (Fmoc)-strategy and its application to the synthesis of cholecytokinin (CCK)-12. Chem. Pharm. Bull. 41, 376–380 (1993).

23. Kitagawa, K., Futaki, S., Yagami, T., Sumi, S. & Inoue, K. Solid-phase synthesis of cionin, a protochordate-derived octapeptide related to the gastrin/cholecystokinin family of peptides, and its mono-tyrosine-sulfate-containing derivatives. Int. J. Pept. Protein Res. 43, 190–200 (1994).

24. Kitagawa, K., Aida, C., Fujiwara, H., Yagami, T. & Futaki, S. Facile solid-phase synthesis of sulfated tyrosine-containing peptides: Part II. Total synthesis of human big gastrin-II and its C-terminal glycine-extended peptide (G34-Gly sulfate) by the solid-phase segment condensation approach. Chem. Pharm. Bull. 49, 958–963 (2001).

25. Koellner, K. M., Smith, M. E. R. & Wong, C. H. Chemoenzymatic synthesis of PSGL-1 glycopeptides: sulfation on tyrosine affects glycosyltransferase-catalyzed synthesis of the O-glycan. Bioorg. Med. Chem. 8, 1017–1025 (2000).

26. Kitagawa, K. et al. Facile solid-phase synthesis of sulfated tyrosine-containing peptides. Total synthesis of human big gastrin-II and cholecystokinin (CCK)-39. J. Org. Chem. 66, 1–10 (2001).

27. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. Nat. Biotechnol. 24, 1436–1440 (2006).

28. Italia, J. S. et al. Genetically encoded protein sulfation in mammalian cells. Nat. Chem. Biol. 16, 379–382 (2020).

29. Kobayashi, T. et al. Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. J. Mol. Biol. 346, 105–117 (2005).

30. Swers, J. S., Kellogg, B. A. & Witrup, K. D. Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. Nucleic Acids Res. 32, e36 (2004).

31. Rosley, A. D. & Stetter, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. Bioimol. Eng. 22, 57–61 (2005).

32. Chin, J. W. et al. An expanded eukaryotic genetic code. Science 301, 964–967 (2003).

33. Chin, J. W., Cropp, T. A., Chu, S., Meggers, E. & Schultz, P. G. Progress toward an expanded eukaryotic genetic code to include unnatural amino acids. Chem. Biol. 10, 511–519 (2003).

34. Lee, H. S., Guo, J., Lemke, E. A., Dimla, R. D. & Schultz, P. G. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in Saccharomyces cerevisiae. J. Am. Chem. Soc. 131, 12921–12923 (2009).

35. Kiga, D. et al. An engineered Escherichia coli tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. Proc. Natl. Acad. Sci. USA 99, 9715–9723 (2002).

36. Lloyd, A. J., Thomann, H.-U., Ibba, M. & Soell, D. A broadly applicable continuous spectrophotometric assay for measuring aminoacyl-tRNA synthetase activity. Nucleic Acids Res. 23, 2886–2892 (1995).

37. Yuan, Z. et al. Controlling multicopy replication of live-attenuated HIV-1 using an unnatural genetic switch. ACS Synth. Biol. 6, 721–731 (2017).

38. Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat. Methods 4, 239–244 (2007).

39. Deiters, A. et al. Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. J. Am. Chem. Soc. 125, 11782–11783 (2003).

40. Veldkamp, C. T. et al. Structural basis of CXCR4 sulfotyrosine recognition by in vivo homologous recombination and yeast surface display. J. Mol. Biol. 253, 2227–2228 (1996).

41. Benatul, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng. Des. Sel. 23, 155–159 (2010).

42. Fong, A. M., Alam, S. M., Imai, T., Haribabu, B. & Patel, D. D. CX3CR1 tyrosine sulfation enhances fractalkine-induced cell adhesion. J. Biol. Chem. 272, 19214–19223 (2007).

43. Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. 24, 1241–1252 (2006).

44. Jørgensen, H. & Farzan, M. Structural basis of CXCR4 sulfotyrosine. Proteins: Struct., Funct. & Genomics 64, 5 (2006).

45. Glynn, D. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345 (2009).

46. Acknowledgements

The authors thank Dr. You Zhou and Ms. Terri Fangman (Morrison Microscopy Core Research Facility at University of Nebraska-Lincoln) for assistance in confocal microscopical analysis. Mr. Dirk Anderson (Flow Cytometry Service Center at University of Nebraska-Lincoln) for assistance in flow cytometry analysis. The authors thank staff at Beamline 12-2 of Stanford Synchrotron Radiation Lightsource for the assistance during the X-ray data collection process. Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences [DE-AC02-76SF00515]. This work was supported by New Faculty Startup Fund from University of Nebraska-Lincoln (to W.N.), National Science Foundation (grant CBT 1805258 to W.N. and J.G.), National Science Foundation (grant MCC 1553041 to J.G.), National Institute of Health (grant 1R01AI111862 to J.G. and W.N.), National Science Foundation (grant CLP 1846908 to J.G. and W.N.). The authors declare no competing interests.

Author contributions

W.N. and J.G. designed the study. X.H., Y.C., and W.N. constructed plasmids. X.H. constructed the EcTyrS library and performed selection in yeast, purified protein and conducted in vitro enzyme assays, expressed and purified protein for crystallography study. X.H. and Y.C. conducted confocal microscopy and flow cytometry studies. J.L. and M.K. performed mass spectrometry study. W.N. conducted protein crystallization and cryo-EM of binned sulfotyrosine, M.K. and E.D. analyzed mass spectrometry data. D.G.B. and L-M.Z. solved macromolecular structure. X.H. and Y.C. conducted conformational analysis, Mr. Dirk Anderson (Flow Cytometry Service Center at University of Nebraska-Lincoln) for assistance in flow cytometry analysis. The authors thanks staff at Beamline 12-2 of Stanford Synchrotron Radiation Lightsource for the assistance during the X-ray data collection process. Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences [DE-AC02-76SF00515]. This work was supported by New Faculty Startup Fund from University of Nebraska-Lincoln (to W.N.), National Science Foundation (grant CBET 1805258 to W.N. and J.G.), National Science Foundation (grant MCC 1553041 to J.G.), National Institute of Health (grant 1R01AI111862 to J.G. and W.N.), National Science Foundation (grant CLP 1846908 to L-M.Z.), National Institute of Health (grant P30 GM103355 to L-M.Z. through the Nebraska Redox Biology Center), and MacNair Scholarship from University of Nebraska-Lincoln (to D.G.B.).

Competing interests

The authors declare no competing interests.
