



Site-specific incorporation of reduction-controlled guest amino acids into proteins for cucurbituril recognition

Zhen Dai^{a,b}, Linzhi Tan^b, Yeyu Su^b, Kerui Zhao^b, Yushun Tian^b, Yu Liu^a, Tao Liu^{b,*}

^a College of Chemistry, State Key Laboratory of Elemento-Organic Chemistry, Nankai University, Tianjin 300071, China

^b State Key Laboratory of Natural and Biomimetic Drugs, Department of Molecular and Cellular Pharmacology, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China

ARTICLE INFO

Article history:

Received 24 June 2023

Revised 9 September 2023

Accepted 17 September 2023

Available online 19 September 2023

Keywords:

Protein recognition

Host-guest recognition

Non-canonical amino acids

Staudinger reduction

Expanded genetic code

ABSTRACT

Protein recognition using host-guest recognition approach is of great interest but has been limited mainly to the protein N-terminal residues. Here, we site-specific incorporated two novel non-canonical amino acids containing supramolecular guest motifs into protein *via* an expanded genetic code. Through Staudinger reduction reactions, the encoded unnatural residues on protein becoming activated and can be specifically recognized by cucurbit[7]uril (CB[7]) and cucurbit[8]uril (CB[8]). We demonstrated that enzyme containing guest amino acid incorporated near the active site can be reversibly regulated by CB[7] recognition, and CB[8] recognition induces protein dimerization. These amino acids will make useful addition to the supramolecular toolbox for protein targeting using molecular recognition approaches.

© 2024 Published by Elsevier B.V. on behalf of Chinese Chemical Society and Institute of Materia Medica, Chinese Academy of Medical Sciences.

Proteins are directly involved in a wide range of cellular activities, and therefore play an important role in cellular processes. The application of supramolecular chemistry to proteins allows reversible manipulation of their properties. Molecular recognition interactions based on host-guest chemistry, which closely resemble protein-protein interactions, have been used to modify, regulate and assemble proteins. It is known that insulin and CB[7] are able to bind efficiently [1], which is mostly attributed to the presence of phenylalanine at the N-terminus of insulin. The positive charge of the N-terminal amino group and the phenyl ring group of phenylalanine causes insulin to have a positively charged aromatic structure. CB[7] can bind positively charged aromatic molecules up to micromolar levels [2]. CB[8] can form ternary complexes with positively charged aromatic molecules. A number of studies have been conducted based on this discovery on homodimeric and heterodimeric proteins derived from cucurbituril [3–6]. However, most of the research focus on proteins with phenylalanine at their N-terminus, which complicates the promotion and application of supramolecular chemistry targeting proteins. Consequently, it is imperative to develop novel guest recognition motifs at the protein level so that site-specific host-guest recognition can be achieved at each site of protein [4,7–9].

Previous studies have searched for natural amino acid sequences with excellent host-guest recognition properties, such

as the "phenylalanine-glycine-glycine" sequence and host-guest recognition of cucurbituril host molecules [10,11]. However, host-guest recognition based on this sequence is restricted to the N-terminal region of the protein [12,13], so there are significant limitations to its application to a broader range of situations. Thus, the development of new technical research methods for studying and exploring new protein sequences for host-guest recognition is highly valuable and potentially applicable.

Genetic code expansion (GCE) involves the introduction of aminoacyl tRNA synthetase that specifically charge non-canonical amino acids to the corresponding orthogonal tRNA in living organisms, such as bacteria or cells [14]. The exogenous orthogonal tRNA recognizes the amber stop codon (UAG) on the mRNA sequence allowing site-directed insertion of non-canonical amino acids into recombinant protein. This facilitates the development of introducing supramolecular chemistry into proteins using building blocks containing guest moieties with high binding affinity to host macrocycles. Previously, we have incorporated tBuF amino acid into proteins site-specifically to reversibly regulate protein function [7]. However, due to the hydrophobic nature and lack of positive charge on tBuF residue, the binding affinity and specificity between tBuF containing protein and CB[7] is moderate, thus limiting its further applications. To improve the system, we hypothesized that amino acid containing both phenyl ring and amine group would be recognized better with CB[7] [15]. However, our effort on the genetic incorporation of such amino acids were unsuccessful, possibly due to the difficulties of genetic encoding

* Corresponding author.

E-mail address: taoliupku@pku.edu.cn (T. Liu).

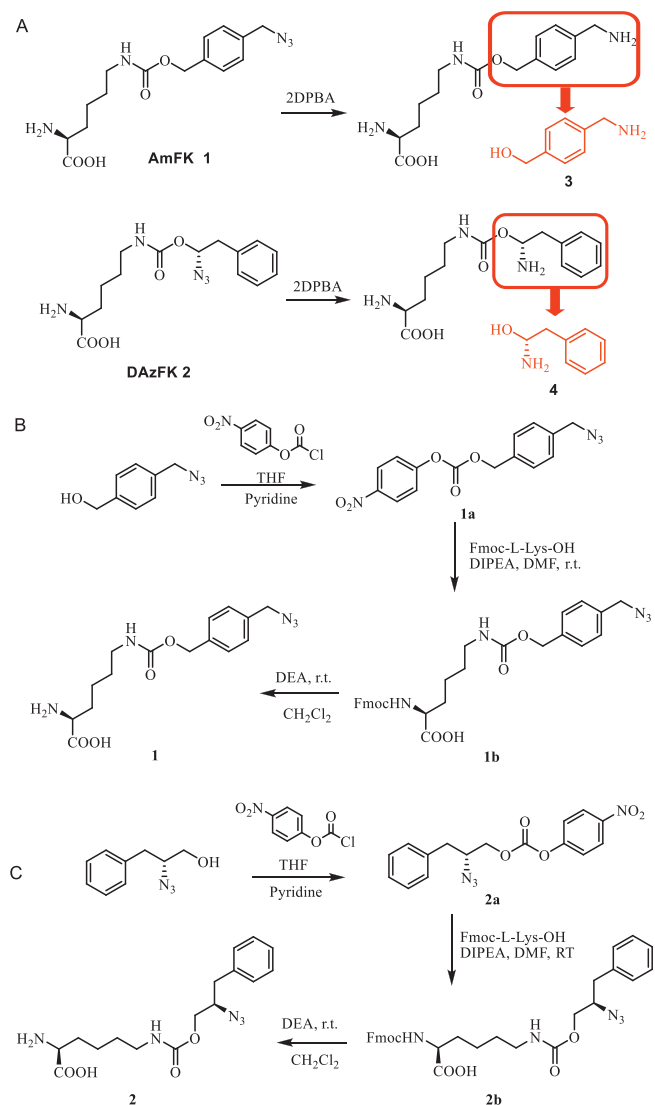


Fig. 1. (A) Synthesis of N^6 -(((4-(azidomethyl)benzyl)oxy)carbonyl)-L-lysine (AmFK, **1**). (B) Synthesis of N^6 -(((*R*)-2-azido-3-phenylpropoxy)carbonyl)-L-lysine (DAzFK, **2**). (C) Reduction of AmFK, guest motif of AmFK: amino methyl)benzyl alcohol **3**. Reduction of DAzFK, guest motif of DAzFK: *D*-phenylalanine alcohol **4**.

of positively charged side chains [16]. To overcome these problems, we have designed and synthesized two novel amino acids N^6 -(((4-(azidomethyl)benzyl)oxy)carbonyl)-L-lysine (AmFK, Fig. 1A, compound **1**) and N^6 -(((*R*)-2-azido-3-phenylpropoxy)carbonyl)-L-lysine (DAzFK, Fig. 1B, compound **2**) with activatable guest motifs (Fig. 1C, compounds **3** and **4**) [17,18], where the azido group on the side chain could be reduced quantitatively with Staudinger reduction with 2DPBA (2-(diphenylphosphino)benzoic acid) [19] to restore the positively charged aromatic structure available for specific CB[7] recognition, the azido specific reduction reaction would also add an additional level of control over the system. Their identities were confirmed by ^1H NMR, ^{13}C NMR and ESI-MS (Figs. S1-S6 in Supporting information).

Firstly, we designed and evaluated the host-guest recognition abilities of the activatable guest motifs of the two non-canonical amino acids (Fig. 1C, compounds **3** and **4**). The affinities of CB[7] binding to compound **3** (4-(amino methyl)benzyl alcohol, Fig. 1C) and compound **4** (*D*-phenylalanine alcohol, Fig. 1C) were measured by ITC and determined to be $3.84\ \mu\text{mol/L}$ and $0.8\ \mu\text{mol/L}$, respectively (Fig. S7 in Supporting information). The complex formation

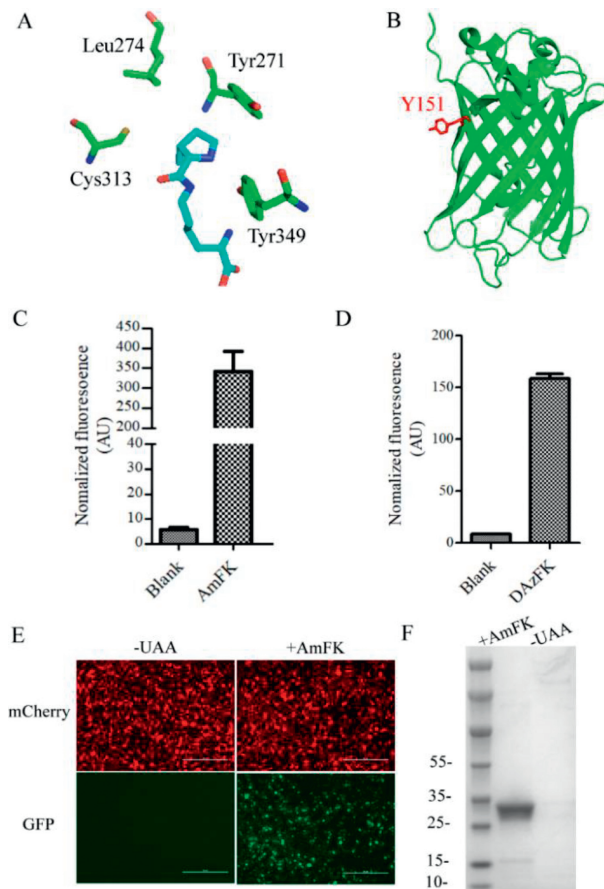


Fig. 2. (A) X-ray crystal structure of the MmPyIRS complex with pyrrolysyl-AMP. The structure is from PDB entry 2ZIM. (B) Crystal structure of sfGFP (green) (PDB ID: 2b3p). Sites selected for incorporation of non-canonical amino acids are indicated in red. (C) AmFK incorporation with ASFRS. (D) DAzFK incorporation with GAVRS. (E) Images of EGFP expressed in the presence of AmFK in mammalian cells. (F) GFP expressed in the presence and absence of 1 mmol/L AmFK with ASFRS.

is then detected by the MALDI-TOF-MS mass spectrometry signal (Fig. S8 in Supporting information). The 1:1 binding mode for the two compounds with CB[7] was further confirmed by measuring the JOB plot with UV-vis spectroscopy (Fig. S9 in Supporting information). After confirming that these two guest moieties could be recognized by and bound to CB[7], we next replaced CB[7] with CB[8] to determine whether it would enable the formation of a ternary complex. ITC experiments on two guest compounds with CB[8] demonstrated that CB[8] had a strong dimerization ability for these guest moieties (Fig. S10 in Supporting information). The JOB plot reveals that CB[8] and two guest moieties are encapsulated at 1:2 stoichiometry (Fig. S11 in Supporting information), demonstrating that dimers can be formed from this positively charged aromatic structure and CB[8].

A large number of structurally diverse UAAs have been genetically encoded in response to an amber codon introduced into a gene of interest using engineered pyrrolysyl tRNA synthetase (PylRS)/tRNA^{CUA} pairs from *Methanosarcina* species [20]. By utilizing PylRS/tRNA^{CUA} pairs and their evolved variants, UAAs have now been incorporated site-specifically into bacteria [21], *Saccharomyces cerevisiae* [14], mammalian cells [22] and living organisms [23]. Screening of a PylRS library designed based on the crystal structure of *M. maize* PylRS pyrrolysyl-AMP (PDB code 2ZIM, Fig. 2A) identified two mutations, named ASFRS (ASFRS:L274A/C313S/Y349F) and GAVRS (GAVRS:Y271G/L274A/C313V/Y349F) for the insertion of these two non-canonical amino acids respectively. Next, sfGFP-

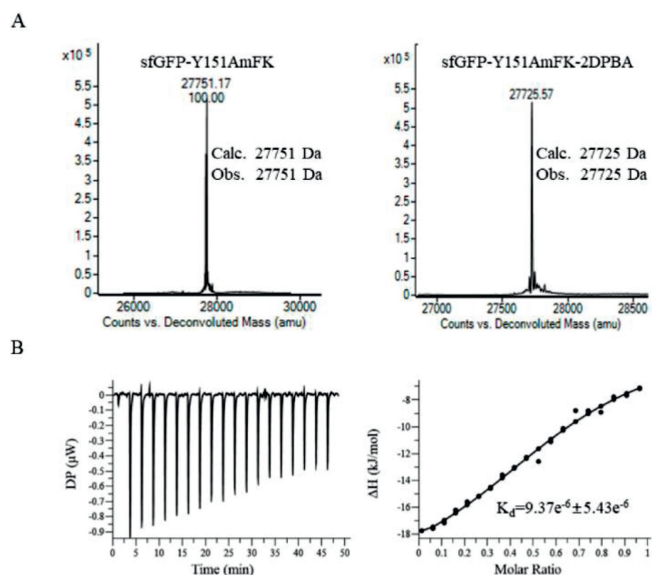


Fig. 3. (A) Treatment of purified sfGFP-Y151AmFK with 2DPBA results in quantitative reduction of the azide group. (B) ITC titration of CB7 into sfGFP-Y151AmFK. After reduction, sfGFP-Y151AmFK-After reduction was at a 50 $\mu\text{mol/L}$ concentration in 10 mmol/L phosphate buffer and CB[7] was at a 250 $\mu\text{mol/L}$ concentration in 10 mmol/L phosphate buffer. The analyses were run at 25 $^{\circ}\text{C}$.

Y151TAG was selected as a model protein to test incorporation efficiency (Fig. 2B). BL21(DE3) cells containing the pET22B-T5-sfGFP-Y151TAG and ASFRS or GAVRS were cultured in the presence or absence of 1 mmol/L ncAAs with the corresponding synthetases. As shown in Figs. 2C and D, a 70 folds and 20 folds fluorescence increase can be observed in the presence of the AmFK and DAZFK, respectively, suggesting an amber codon suppression of both ncAAs. As shown in Fig. 2E, the full-length GFP can only be observed in the presence of 1 mmol/L AmFK. Additionally, the incorporation of the AmFK in mammalian cells was also evaluated by co-transfection of pcmv-ASFRS with pcDNA3.1-mCherry-EGFP-39TAG (Fig. 2F).

To verify the incorporation of ncAAs, the resulting protein mutants were purified by Ni-NTA affinity chromatography, and the purified sfGFP can be obtained at yields up to 60 mg/L of cell culture. The identity of these ncAAs on sfGFP were confirmed by high-resolution mass spectrometry [quadrupole time-of-flight/MS (QTOF/MS)]. The mass spectrometric traces showed that AmFK-sfGFP and wild-type sfGFP gave an observed mass of 27751 Da and 27597 Da, respectively (Fig. S12 in Supporting information). The mass difference of 154 Da is consistent with the molecular difference between tyrosine and AmFK confirming the genetic incorporation of AmFK into sfGFP. On the other hand, the DAZFK-sfGFP shows an observed mass of 27765 Da, which is consistent with the incorporation of DAZFK (Fig. S13 in Supporting information). Since the DAZFK is synthetically more difficult than that of AmFK, AmFK was preferred to use in the following experiments due to its higher incorporation efficiency and lower synthetic expense.

Next, we determined if the AmFK sidechain can be activated on the protein. Mutant AmFK-sfGFP was reduced with 2 equiv. of 2DPBA at room temperature. The quantitative reduction of AmFK residue on sfGFP was determined by ESI-MS analysis. As shown in Fig. 3A, the observed mass after reduction matches with the calculated mass with mass difference of -26 Da. We further confirm that the reduction not only works at the purified protein level but also in cell lysate (Fig. S14 in Supporting information). Furthermore, we conducted the same experiment with DAZFK-sfGFP and obtained similar results (Fig. S13). These data suggested that AmFK and DAZFK can be site-specifically introduced into protein

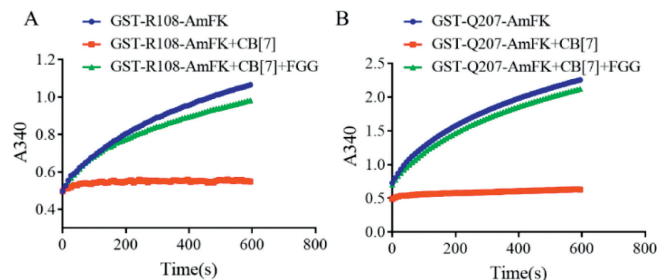


Fig. 4. Assay of enzymatic activity of the GST-Q207AmFK mutant (A) and R108AmFK mutant (B) (pH 6.5, phosphate-buffered saline, 25 $^{\circ}\text{C}$, [GST-R108-AmFK] = 3 $\mu\text{mol/L}$, [GST-Q207-AmFK] = 3 $\mu\text{mol/L}$, [CB[7]] = 1 mmol/L, [FGG] = 1.2 mmol/L, [GSH] = 1 mmol/L, [CDNB] = 1 mmol/L (CDNB = 1-chloro-2,4-dinitrobenzene)).

via an expanded genetic code, and can be reduced quantitatively on the resulting protein mutants to afford positively charged aromatic guest moieties available for further host molecular recognition.

Next, we evaluated the host-guest molecular recognition on these sfGFP-151 mutants with CB[7]. Following the reduction of the azide, microthermophoresis dynamometer was applied to determine the binding affinity of CB[7] and sfGFP mutants. The K_d values for the binding of CB[7] to sfGFP-Y151AmFK and sfGFP-Y151DAzFK were determined to be 1.22 $\mu\text{mol/L}$ and 0.72 $\mu\text{mol/L}$, respectively (Fig. S15 and Table S1 in Supporting information), whereas no binding was observed between wild sfGFP-WT and CB[7]. These data were consistent with the binding data obtained by ITC experiments (Fig. 3B, Fig. S16 in Supporting information), suggesting that the observed binding is indeed a result of specific molecular recognition between the host molecule CB[7] and the reduced guest motifs encoded by AmFK and DAZFK.

Next we determine whether our host-guest molecular recognition system can reversible controlling enzyme activity, we applied an assay using glutathione S-transferase (GST) as a model system [7]. Two GST mutants, GST-Q108tBuF and GST-Q207AmFK, were recombinantly expressed and purified. The activities of the mutant enzymes were measured in the presence of 1 mmol/L CB[7]. A near complete inhibition of the enzymatic activity for both mutants were observed, and addition of only 1.2 equiv. of FGG (1.2 mmol/L) as a competitive guest could restore the activity of both enzymes (Figs. 4A and B). It worth to mention that the inhibition result on Q108 position using reduced AmFK is better than tBuF, a guest amino acid used in our previous study [7]. These results suggested that AmFK residue on protein is a reduction controlled, more effective guest recognition motif, which can achieve single-residue resolution supramolecular host-guest recognition and binding to efficiently and reversibly regulate enzyme function.

Next, we investigated whether reduced AmFK residue can be recognized by CB[8] to induce protein dimerization. Reduced sfGFP-Y151AmFK were incubated with CB[8], and dimer formation was analyzed by analytical ultracentrifugation (AUC) [24]. As shown in Fig. 5 and Fig. S17 (Supporting information), when host macrocyclic molecules CB[8] were added, there was an increase in sfGFP dimerization formation up to 14%. This results confirm that the activatable guest motif can be recognized and bound by CB[8], leading to the formation of protein dimers in a host dependent manner.

Overall, we have genetically encoded two reduction-controlled guest amino acids, AmFK and DAZFK, site-specifically in proteins via an expanded genetic code. Supramolecular host-guest interactions on the resulting protein mutants allowing a reversible control of protein activity and protein assembly. The reduced residues have high affinity and specificity towards both CB[7] and CB[8], there-

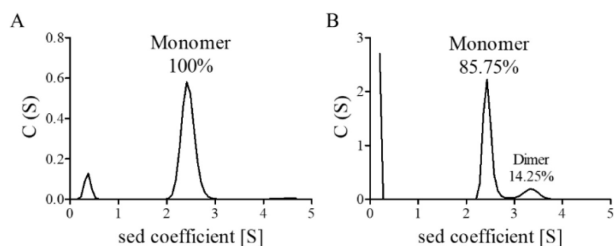


Fig. 5. AUC assay of enzymatic activity of the sfGFP-AmFK protein dimerisation (pH 7.0, phosphate-buffered saline, 25 °C). (A) 1 μ mol/L sfGFP-AmFK-AR; (B) 10 μ mol/L sfGFP-AmFK-AR and 10 μ mol/L CB[8]. AR: after reduction.

fore expected to be valuable for the study of protein function and regulation.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Nos. 22325701, U22A20332, 92156025 and 92253301), the National Key Research and Development Program of China (Nos. 2022YFA0912400 and 2021YFA0909900), and the Beijing Natural Science Foundation (No. JQ20034). We thank Prof. Weimin Xuan from the Tianjin University for assistance with the experiments. We also thank Dr. Qian Wang at the SKLNBD Research for assistance in ITC experiments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ccl.2023.109121.

References

- [1] M.J. Webber, E.A. Appel, B. Vinciguerra, et al., *Proc. Nat. Acad. Sci. U. S. A.* 113 (2016) 14189–14194.
- [2] A.A. Smith, C.L. Maikawa, G.A. Roth, E.A. Appel, *Org. Biomol. Chem.* 18 (2020) 4371–4375.
- [3] P.J. de Vink, J.M. Briels, T. Schrader, et al., *Angew. Chem. Int. Ed.* 56 (2017) 8998–9002.
- [4] Y.M. Zhang, J.H. Liu, Q. Yu, X. Wen, Y. Liu, *Angew. Chem. Int. Ed.* 58 (2019) 10553–10557.
- [5] C. Hou, Z. Huang, Y. Fang, J. Liu, *Org. Biomol. Chem.* 15 (2017) 4272–4281.
- [6] X. Yang, R. Wang, A. Kermagoret, D. Bardelang, *Angew. Chem. Int. Ed.* 59 (2020) 21280–21292.
- [7] W. Cao, X. Qin, Y. Wang, et al., *Angew. Chem. Int. Ed.* 60 (2021) 11196–11200.
- [8] Y.L. Ma, C. Sun, Z. Li, et al., *CCS Chem.* 4 (2022) 1977–1989.
- [9] L. Yue, K. Yang, J. Wei, et al., *CCS Chem.* 4 (2022) 1745–1757.
- [10] S.J. Barrow, S. Kaseira, M.J. Rowland, J. Del Barrio, O.A. Scherman, *Chem. Rev.* 115 (2015) 12320–12406.
- [11] C.L. Maikawa, A.A.A. Smith, L. Zou, et al., *Nat. Biomed. Eng.* 4 (2020) 507–517.
- [12] D.T. Dang, J. Schill, L. Brunsveld, *Chem. Sci.* 3 (2012) 2679–2684.
- [13] C. Hou, J. Li, L. Zhao, et al., *Angew. Chem. Int. Ed.* 52 (2013) 5590–5593.
- [14] C.C. Liu, P.G. Schultz, *Annu. Rev. Biochem.* 79 (2010) 413–444.
- [15] L.A. Logsdon, C.L. Schardon, V. Ramalingam, S.K. Kwee, A.R. Urbach, *J. Am. Chem. Soc.* 133 (2011) 17087–17092.
- [16] D.P. Nguyen, P.B. Kapadnis, H. Neumann, J.W. Chin, *J. Am. Chem. Soc.* 131 (2009) 14194–14195.
- [17] A.L.Z. Lee, Z.X. Voo, W. Chin, et al., *ACS Appl. Mater. Interfaces* 10 (2018) 13274–13282.
- [18] J.F. Jensen, K. Worm-Leonhard, M. Meldal, *Eur. J. Org. Chem.* 2008 (2008) 3785–3797.
- [19] M. Fottner, A.D. Brunner, V. Bittl, et al., *Nat. Chem. Biol.* 15 (2019) 276–284.
- [20] Y. Huang, T. Liu, *Sig. Transduct. Target. Ther.* 5 (2020) 155–156.
- [21] J. Li, S. Lin, J. Wang, et al., *J. Am. Chem. Soc.* 135 (2013) 7330–7338.
- [22] A. Chatterjee, H. Xiao, M. Bollong, H.W. Ai, P.G. Schultz, *Proc. Nat. Acad. Sci. U. S. A.* 110 (2013) 11803–11808.
- [23] S. Greiss, J.W. Chin, *J. Am. Chem. Soc.* 133 (2011) 14196–14199.
- [24] G.B. Edwards, U.M. Muthurajan, S. Bowerman, K. Luger, *Curr. Protoc. Mol. Biol.* 133 (2020) e131–e157.