Aptamer-Array-Guided Protein Assembly Enhances Synthetic mRNA Switch Performance

Qiuyu Lu, Yaxin Hu, Cheuk Yin Li, and Yi Kuang*

Abstract: Synthetic messenger RNA (mRNA) switches are powerful synthetic biological tools that can sense cellular molecules to manipulate cell fate. However, their performances are limited by high output signal noise due to leaky output protein expression. Here, we designed a readout control module that disables protein leakage from generating signal. Aptamer array on the switch guides the inactive output protein to self-assemble into functional assemblies that generate output signal. Leaky protein expression fails to saturate the array, thus produces marginal signal. In this study, we demonstrated that switches with this module exhibit substantially lower signal noise and, consequently, higher input sensitivity and wider output range. Such switches are applicable for different types of input molecules and output proteins. The work here demonstrates a new type of spatially guided protein self-assembly, affording novel synthetic mRNA switches that promise accurate cell manipulation for biomedical applications.

Synthetic mRNA switches are important type of artificial circuits that promise programmed control of living cells. Synthetic mRNA switches sense cellular molecules or traits (input) to change the expression level of the encoded protein, producing output signals for monitoring or manipulating of the target cells.[1] Because of their high biocompatibility, biodegradability and no interference with the genomes, synthetic mRNA switches promise safe cell-based solutions for biomedical problems in clinical settings, such as cellular therapeutics, tissue engineering and regenerative medicine.[2]

[*] Q. Lu, Y. Hu, C. Yin Li, Prof. Y. Kuang
Department of Chemical and Biological Engineering, The Hong
Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong (Hong Kong)
E-mail: kekuang@ust.hk

Prof. Y. Kuang
HKUST Shenzhen Research Institute
Shenzhen, Guangdong (China)

Synthetic mRNA switches sense cellular input via non-covalent bonds (mostly hydrogen bonds).[3] For example, the most well-studied microRNA (miRNA) sensing switch depends on hydrogen bonding formation between miRNA and the switch to trigger output signal repression.[4] Because noncovalent bonds are generally weak, the limited input-switch binding affinities yield poor output performance, especially high signal noise due to leaky output expression in the presence of input.[5] For example, Saito et al. discovered that miRNA sensing switches can only generate low-to-high output signal change upon on-to-off input transition.[6] Several advancements have attempted to enhance the binding affinities, including increasing the number of input binding sites,[7] optimizing input binding sequences,[8] and using thermal stable modified nucleotides.[9] However, as the underlying weak binding remains the same, the high signal noise continued to limit the performance of current synthetic mRNA switches.[10]
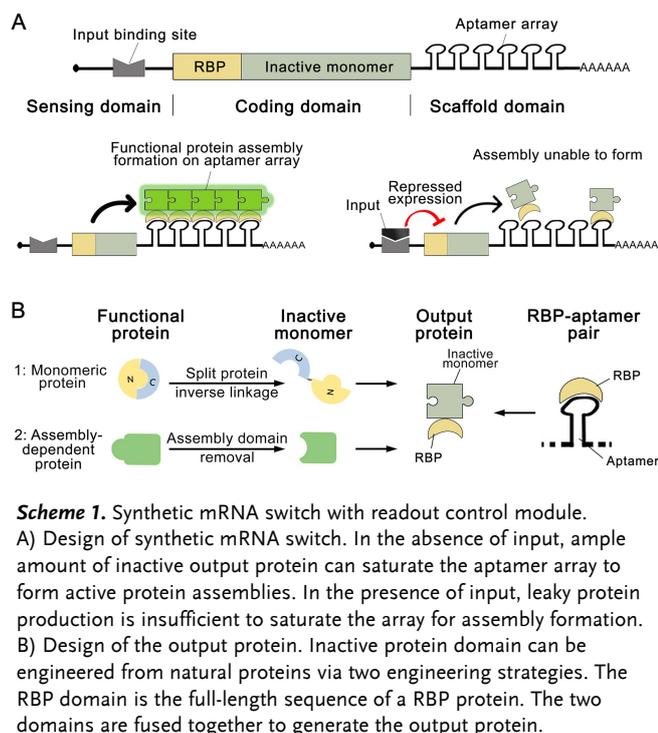
Harnessing the principles of supramolecular chemistry, we engineered a readout control module to circumvent the output signal leakage issue of synthetic mRNA switches. The readout control module consists of inactive output proteins and an aptamer array. The aptamer array serves as scaffolding to enrich the inactive proteins to close proximity. When the inactive proteins saturate the array, self-assembly among the proteins affords functional assemblies. Thus, without altering the input-switch binding, the module creates a concentration threshold that blocks the protein expression leakage from producing signal. We demonstrated that this readout control module can successfully enrich output protein in cellular environment and can drastically boost the performance of various synthetic mRNA switches, significantly reducing signal noise, increasing input sensitivity and output operation range. (i.e., the difference between maximum and minimum signals). We also demonstrated that fluorescent proteins and enzymes can both be engineered to serve as the output protein of the switches, promising versatile downstream cell manipulation with high accuracy.

To construct the inactive output protein that can be recognized by the aptamer array, a well-studied RNA binding protein (RBP) MS2 coat protein (MCP) was used as the binding domain (Scheme 1). The functional domain of the output protein must be inactive in monomer form and active in a higher-order assembly. To achieve this purpose, we utilized a protein engineering strategy to inversely link split protein fragments of a natural protein. While such inverse linkage prevents the fragments from intramolecular association, the resulted protein is inactive.[11] The inversely

Scheme 1. Synthetic mRNA switch with readout control module. A) Design of synthetic mRNA switch. In the absence of input, ample amount of inactive output protein can saturate the aptamer array to form active protein assemblies. In the presence of input, leaky protein production is insufficient to saturate the array for assembly formation. B) Design of the output protein. Inactive protein domain can be engineered from natural proteins via two engineering strategies. The RBP domain is the full-length sequence of a RBP protein. The two domains are fused together to generate the output protein.

Figure 1. Engineering of model readout control module. A) Structures and basal fluorescence levels of the engineered proteins. B) Effect of the MS2 aptamer array linker length on the MCP-Inverted sfGFP assembly formation. C) Time curves of GFP signal production from HEK293 cells transfected with sfGFP or sfGFP-Array mRNAs. D) Representative confocal images showing the spatial distribution of fluorescent signals in HEK293 cells transfected with sfGFP or sfGFP-Array mRNAs. N = 3; data are presented as the mean ± SD. **$P < 0.01$ calculated through single-factor analysis of variance (ANOVA).

linked super-fold green fluorescent protein (denoted as Inverted sfGFP) was used as the model functional domain, producing a model output protein denoted as MCP-Inverted sfGFP. As shown in Figure 1A, we evaluated the basal GFP signal produced by mRNAs encoding different proteins on HEK293 cells. Only marginal GFP signal was detected from the MCP-Inverted sfGFP mRNA, which demonstrated that the bulky MCP domain can prevent the Inverted sfGFP domain from random assembly.

Next, to construct the aptamer array as the output protein assembly scaffolding, we used 12 copies of MS2 aptamer, which exhibits a strong binding affinity with MCP.[12] Linker sequences used in previous works (10, 20 and 51 nucleotides [nt])[13] were tested. As shown in Figure 1B, the mRNA carrying array with 51 nt linker corresponded to the strongest GFP signal, which as comparable to that of the mRNA encoding native sfGFP (Figure S1A). The 51 nt linker in the minimum free energy structure has an estimated length around 15 nm. Together with its flexibility, output proteins of difference sizes can be hosted by the 12×51 array for assembly. Thus, the 12×51 array was used in the following experiment.

We then examined the duration of the output signal production from sfGFP-Array mRNA or sfGFP mRNA on HEK293 cells. As shown in Figure 1C, a slightly delayed signal peak arrival and narrower peak width was observed from sfGFP-Array mRNA, this could be attributed to the additional steps of spatial organization and assembly formation for signal production. Nonetheless, the time curves showed that sfGFP-Array mRNA can generate a sufficient output signal observation window. Interestingly, the half-life of transfected mRNA has been reported to be approximately 12 hours after transfection.[9b] The prolonged GFP
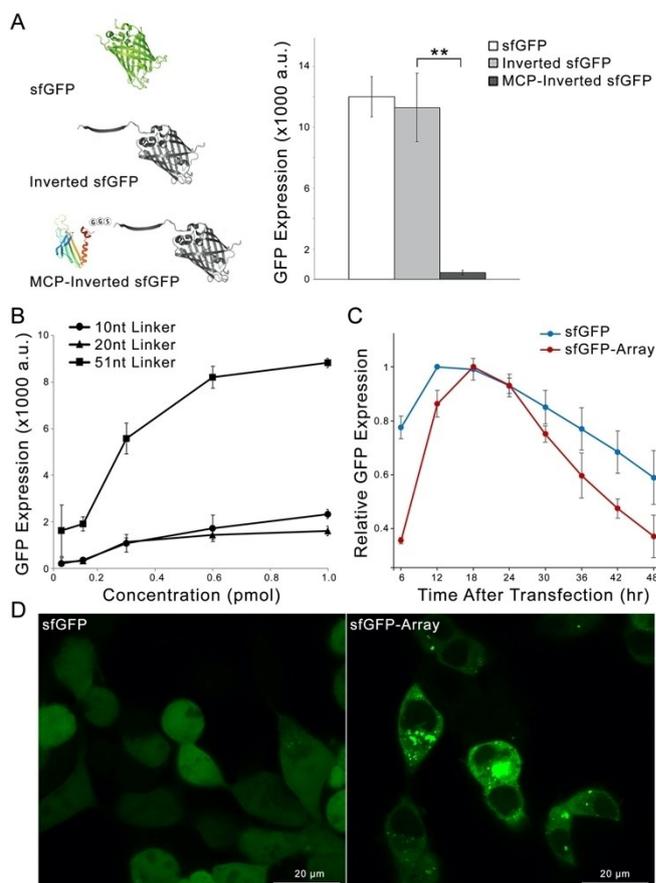
signal observation may be attributable to that the protein assemblies can maintain their oligomeric form after dissociation from the scaffolding,[14] as the self-assembly was dependent on the binding between the split protein fragments. This data confirmed that the RBP-inversely-linked protein and aptamer array represented a suitable readout control module for synthetic mRNA switch.

Next, we studied the spatial distribution of GFP signal from HEK293 cells. As shown in Figure 1D (see also Figure S1D), the GFP signal produced from sfGFP mRNA spread all over the cell, as sfGFP proteins can freely diffuse across nuclear pores. However, the GFP signal produced from sfGFP-Array mRNA exhibited mostly as strong dots in the cytosol, a pattern that strongly associated GFP signal with the localization of the mRNA. Electrophoretic mobility shift assay on 12×51 aptamer array RNA also showed clear upward shift when incubated with both MCP and MCP-Inverted sfGFP proteins, suggesting strong binding between the aptamer array and the proteins (Figure S2A). These

data together strongly implied that the aptamer array is essential for generating the output signal.

To evaluate the impact of readout control module on synthetic mRNA switches, we applied the module onto the well-known miRNA-21-sensing mRNA switch (denoted as miR-21-sfGFP-Array switch), which carries anti-miRNA-21 sequences on the 5′-UTR. Previous work by Miki and co-workers demonstrated that such design offers strong miRNA sensitivity.[15] As shown in Figure 2A, a classic switch encoding an intact sfGFP was used as reference (denoted as miR-21-sfGFP switch). HEK293 cells were co-transfected with the switches and a gradient concentration of miR-21 mimic. As presented in Figure 2B, the miR-21-sfGFP-Array switch exhibited stronger miRNA sensitivity, with the IC_{50} value being more than an order of magnitude smaller than that of the miR-21-sfGFP switch. The miR-21-sfGFP switch still produced more than 50% of the output signal when the mimic:switch ratio reached 20 (3 pmol of mimic), whereas the output signal for the miR-21-sfGFP-Array switch dropped to near background level (Figure S3A).

We then examined the function of the miR-21-sfGFP-Array switch for detecting endogenous miRNA-21. A pair of cells with differential endogenous miR-21 levels, HEK293 and HepG2-RFP cells (HepG2 knocked in with a red fluorescent protein gene), was used.[16] The two switches were transfected to the co-cultured cells. sfGFP mRNA and
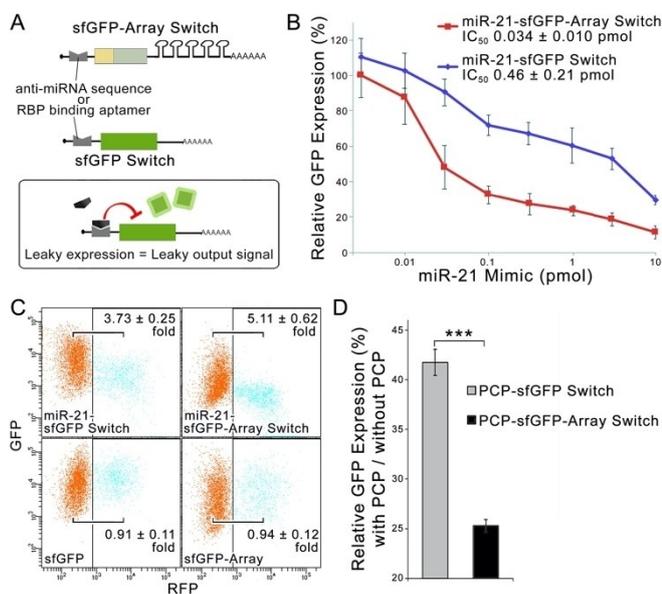
sfGFP-Array mRNA were also transfected to reveal the basal signal levels in the two cells. As shown in Figure 2C, the miR-21-sfGFP-Array switch clearly induced a stronger GFP signal difference between the two populations than the miR-21-sfGFP switch. Moreover, the GFP level from miR-21-sfGFP-Array switch in the miRNA positive HepG2-RFP cells was at a comparable level as that in the negative control (Figure S4), suggesting nearly full output signal noise repression.

We also employed the readout control module on a protein sensing switch. Module input protein PP7 coat protein (PCP) was artificially introduced via co-transfection of PCP mRNA. The GFP signals from the conventional switch (denoted as PCP-sfGFP switch) and a switch with the readout control module (denoted as PCP-sfGFP-Array switch) without PCP mRNA co-transfection were set as 100%, respectively. As shown in Figure 2D, the PCP-sfGFP-Array switch exhibited significant signal repression with PCP mRNA co-transfection. This substantially stronger GFP signal reduction confirmed that the readout control module can be applied to synthetic mRNA switches sensing different inputs.

As shown in Figure 3A, we explored an alternative protein engineering strategy to employ a well-known assembly-dependent mouse Caspase 8 protein (denoted as Caspase 8). When mouse Caspase 8 proteins were highly expressed in human cell line, they randomly assemble in the lack of upstream signal to induce weak cytotoxicity.[8b] MCP domain replaced the death effector domain (DED), producing the Caspase 8-Array mRNA (carrying MCP-dCaspase 8 and 12×51 array). Figure 3B shows that Caspase 8-Array mRNA induced the strongest cytotoxicity, with severe drop

Figure 2. The readout control module enhances the performance of model mRNA switches. A) Constructs of the mRNA switches. Inset: scheme showing the output signal noise generated from the reference switch. B) The miRNA sensitivity of the pair of miR-21-sensing switches. C) Representative dot-plot images showing the cell-type identification performance of the pair of miR-21 sensing switches. Orange dots indicate the miR-21 negative HEK293 cells; blue dots indicate the miR-21 positive HepG2 cells. D) The performance of a pair of PCP-sensing switches. N = 3; data are presented as the mean ± SD. ***P < 0.001 calculated through single-factor analysis of variance (ANOVA).
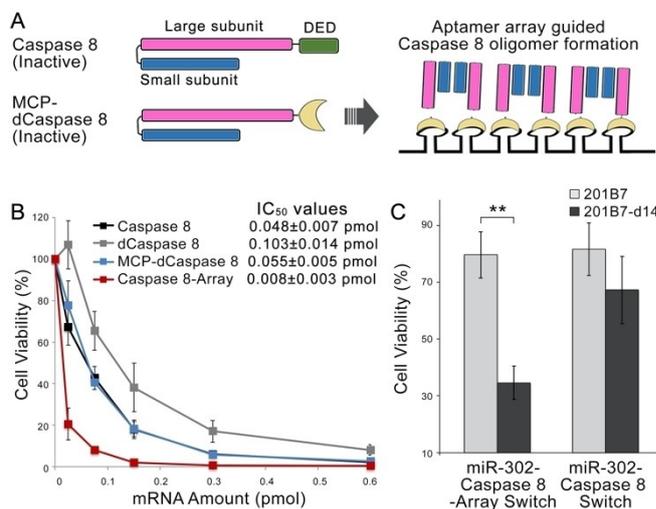

Figure 3. Engineering of mRNA switch with assembly-deficient proteins as the inactive protein domain. A) Schemes of mouse Caspase 8 and the according engineered proteins. After self-assembly on the aptamer array, the Caspase 8 assemblies self-catalyze to induce downstream cytotoxicity. B) Cytotoxicity of mRNAs encoding different engineered proteins. C) Cell-specific elimination performance by the pair of switches. N = 3; data are presented as the mean ± SD. **P < 0.01 calculated through single-factor analysis of variance (ANOVA).

in viability observed even at lowest concentration of mRNA used. This result suggested that the aptamer array enrich and drive cytotoxic self-assembly formation.

Next, we constructed a miR-302 sensing Caspase 8-Array switch (denoted as miR-302-Caspase 8-Array switch). miR-302 is a cell stemness marker that has been used to identify induced pluripotent stem cells (iPSCs) from their differentiated ones.[17] iPSC cell line 201B7 and its differentiated cell line, 201B7-d14, were transfected with the miR-302-Caspase 8-Array switch and reference miR-302-Caspase 8 switch (encoding MCP-dCaspase 8). As shown in Figure 3C, the miR-302-Caspase 8-Array switch induced clear difference in cytotoxicity on the two cells. The miR-302-Caspase 8 switch exhibited weak cytotoxicity even on the miRNA negative cells, which was likely due to low basal protein expression level. We altered the natural m7G(5′)ppp(5′)G (m7G) cap into synthetic cap analogs to tune the basal protein expression level (Figure S5C). Regardless of the cap analogs used, the miR-302-Caspase 8-Array switch consistently showed stronger cytotoxicity difference on the two cells. These results indicated that assembly-dependent proteins can also be engineered to serve to output proteins for the novel mRNA switch.

This research provides a new module that can spatially localize and self-assemble proteins. It is a pioneer design that the mRNA not only produces but also controls the behavior of its encoding protein. As many key cellular proteins function only at higher-order assembly form,[18] the readout control module in this paper highlights a promising new direction for engineering functional supramolecular assemblies[19] and nanostructures in living cells.[20] On the other hand, aptamer arrays have long been used to attract reporter proteins or chemical dyes for RNA labelling.[21] A new application of the aptamer arrays, that is to control the function of a single protein, is presented in this work. Thus, the readout control module expands the functional elements in synthetic mRNA toolkit, facilitating the future engineering of other complex synthetic biology tools.

The synthetic mRNA switches with the readout control modules all exhibited substantially enhanced output performance, featured with the near-complete sequestering of output signal noise, stronger input sensitivity, and a wider output operation range (e.g., 90% in signal level on the miR-21-sensing switch). We found that such advanced synthetic mRNA switches can sense different types of input molecules, host different types of output protein, and, importantly, use different RBP-aptamer pairs to construct the readout control module (Figure S6). These properties enable the further engineering of complex RNA circuits that offer accurate and versatile orthogonal controls of cell fate. However, a limitation lies in that the proteins that function in the organelle or outside of the cell cannot be employed. Further engineering of mRNA switches that can control the readout of these proteins should be explored. In conclusion, we believe that the advanced synthetic mRNA switch with readout control module enables efficient and accurate cell fate manipulation, thereby promising the biomedical applications of synthetic mRNA switches for clinical purposes.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the Supporting Information of this article.

[1] a) H. Saito, T. Kobayashi, T. Hara, Y. Fujita, K. Hayashi, R. Furushima, T. Inoue, *Nat. Chem. Biol.* **2010**, *6*, 71–78; b) S. J. Culler, K. G. Hoff, C. D. Smolke, *Science* **2010**, *330*, 1251–1255; c) L. Warren, P. D. Manos, T. Ahfeldt, Y. H. Loh, H. Li, F. Lau, W. Ebina, P. K. Mandal, Z. D. Smith, A. Meissner, G. Q. Daley, A. S. Brack, J. J. Collins, C. Cowan, T. M. Schlaeger, D. J. Rossi, *Cell Stem Cell* **2010**, *7*, 618–630; d) C. I. An, V. B. Trinh, Y. Yokobayashi, *RNA* **2006**, *12*, 710–716; e) K. C. Venkata Subbaiah, O. Hedaya, J. Wu, F. Jiang, P. Yao, *Comput. Struct. Biotechnol. J.* **2019**, *17*, 1326–1338.

[2] a) L. Wroblewska, T. Kitada, K. Endo, V. Siciliano, B. Stillo, H. Saito, R. Weiss, *Nat. Biotechnol.* **2015**, *33*, 839–841; b) Y. Y. Chen, M. C. Jensen, C. D. Smolke, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8531–8536.

[3] Y. Yokobayashi, *Curr. Opin. Chem. Biol.* **2019**, *52*, 72–78.

[4] U. Schmitz, X. Lai, F. Winter, O. Wolkenhauer, J. Vera, S. K. Gupta, *Nucleic Acids Res.* **2014**, *42*, 7539–7552.

[5] R. Kent, N. Dixon, *ACS Synth. Biol.* **2019**, *8*, 884–901.

[6] H. Saito, Y. Fujita, S. Kashida, K. Hayashi, T. Inoue, *Nat. Commun.* **2011**, *2*, 160.

[7] M. Hirosawa, Y. Fujita, C. J. C. Parr, K. Hayashi, S. Kashida, A. Hotta, K. Woltjen, H. Saito, *Nucleic Acids Res.* **2017**, *45*, e118.

[8] a) H. Nakanishi, *Life* **2021**, *11*, 1192; b) T. Shibata, Y. Fujita, H. Ohno, Y. Suzuki, K. Hayashi, K. R. Komatsu, S. Kawasaki, K. Hidaka, S. Yonehara, H. Sugiyama, M. Endo, H. Saito, *Nat. Commun.* **2017**, *8*, 540; c) J. Yang, S. Ding, *Mol. Ther. Nucleic Acids* **2021**, *25*, 683–695.

[9] a) J. Lockhart, J. Canfield, E. F. Mong, J. VanWye, H. Totary-Jain, *Mol. Ther. Nucleic Acids* **2019**, *14*, 339–350; b) C. J. C. Parr, S. Wada, K. Kotake, S. Kameda, S. Matsuura, S. Sakashita, S. Park, H. Sugiyama, Y. Kuang, H. Saito, *Nucleic Acids Res.* **2020**, *48*, e35.

[10] M. Finke, D. Brecht, J. Stifel, K. Gense, M. Gamerdinger, J. S. Hartig, *Nucleic Acids Res.* **2021**, *49*, e71.

[11] Y. E. Kim, Y. Kim, J. A. Kim, H. M. Kim, Y. Jung, *Nat. Commun.* **2015**, *6*, 7134.

[12] a) W. T. Horn, M. A. Convery, N. J. Stonehouse, C. J. Adams, L. Liljas, S. E. V. Phillips, P. G. Stockley, *RNA* **2004**, *10*, 1776–1782; b) S. Rowsell, N. J. Stonehouse, M. A. Convery, C. J. Adams, A. D. Ellington, I. Hirao, D. S. Peabody, P. G. Stockley, S. E. V. Phillips, *Nat. Struct. Biol.* **1998**, *5*, 970–975.

[13] a) E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, R. M. Long, *Mol. Cell* **1998**, *2*, 437–445; b) B. Wu, J. A. Chao, R. H. Singer, *Biophys. J.* **2012**, *102*, 2936–2944; c) B. Wu, J. Chen, R. H. Singer, *Sci. Rep.* **2014**, *4*, 3615.

[14] a) K. Do, S. G. Boxer, *J. Am. Chem. Soc.* **2011**, *133*, 18078–18081; b) M. G. Romei, S. G. Boxer, *Annu. Rev. Biophys.* **2019**, *48*, 19–44.

[15] K. Miki, K. Endo, S. Takahashi, S. Funakoshi, I. Takei, S. Katayama, T. Toyoda, M. Kotaka, T. Takaki, M. Umeda, C. Okubo, M. Nishikawa, A. Oishi, M. Narita, I. Miyashita, K. Asano, K. Hayashi, K. Osafune, S. Yamanaka, H. Saito, Y. Yoshida, *Cell Stem Cell* **2015**, *16*, 699–711.

[16] a) S. Zhu, H. Wu, F. Wu, D. Nie, S. Sheng, Y. Mo, *Cell Res.* **2008**, *18*, 350–359; b) B. Mao, B. Xiao, Z. Zhang, D. Wang, G. Wang, *Mol. Med.* **2015**, *12*, 4917–4924.

[17] C. J. C. Parr, S. Katayama, K. Miki, Y. Kuang, Y. Yoshida, A. Morizane, J. Takahashi, S. Yamanaka, H. Saito, *Sci. Rep.* **2016**, *6*, 32532.

[18] a) J. J. McManus, P. Charbonneau, E. Zaccarelli, N. Asherie, *Curr. Opin. Colloid Interface Sci.* **2016**, *22*, 73–79; b) V. M. Bolanos-Garcia, Q. Wu, T. Ochi, D. Y. Chirgadze, B. L. Sibanda, T. L. Blundell, *Philos. Trans. R. Soc. London Ser. A* **2012**, *370*, 3023–3039.

[19] a) W. Tan, Q. Zhang, J. Wang, M. Yi, H. He, B. Xu, *Angew. Chem. Int. Ed.* **2021**, *60*, 12796–12801; *Angew. Chem.* **2021**, *133*, 12906–12911; b) Y. Yuan, L. Wang, W. Du, Z. Ding, J. Zhang, T. Han, L. An, H. Zhang, G. Liang, *Angew. Chem. Int. Ed.* **2015**, *54*, 9700–9704; *Angew. Chem.* **2015**, *127*, 9836–9840.

[20] a) L. Yang, A. Liu, S. Cao, R. M. Putri, P. Jonkheijm, J. J. L. M. Cornelissen, *Chem. Eur. J.* **2016**, *22*, 15570–15582; b) D. T. Seroski, G. A. Hudalla in *Biomedical Applications of Functionalized Nanomaterials*, (Eds.: B. Sarmento, J. das Neves), Elsevier, Amsterdam, **2018**, pp. 569–598.

[21] a) C. J. Delebecque, P. A. Silver, A. B. Lindner, *Nat. Protoc.* **2012**, *7*, 1797–1807; b) D. Gemmill, S. D'souza, V. Meier-Stephenson, T. R. Patel, *Biochem. Cell Biol.* **2020**, *98*, 31–41; c) T. Wang, J. Ray, *Protein Cell* **2012**, *3*, 739–754; d) M. Panigaj, M. B. Johnson, W. Ke, J. McMillan, E. A. Goncharova, M. Chandler, K. A. Afonin, *ACS Nano* **2019**, *13*, 12301–12321.