Optics-Free Visualization of Proteins in Single Cells by Time of Flight-Secondary Ion Mass Spectrometry Coupled with Genetically Encoded Chemical Tags

Feifei Jia[a], Jie Wang[a,b], Yanyan Zhang[a], Qun Luo[a,c], Luyu Qi[a,c], Yinzhu Hou[a,c], Jun Du[b], Yao Zhao[a*] and Fuyi Wang[a,c,d*]

[a] Beijing National Laboratory for Molecular Sciences; National Centre for Mass Spectrometry in Beijing; CAS Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, P. R. China.
[b] College of Chemistry and Materials Science, Key Laboratory of Functional Molecular Solids, the Ministry of Education, Anhui Laboratory of Molecular-Based Materials, Anhui Normal University, Wuhu, 241000, P. R. China
[c] University of Chinese Academy of Sciences, Beijing 100049, P. R. China.
[d] Basic Medical College, Shandong University of Chinese Traditional Medicine, Jinan 250355, P. R. China

ABSTRACT: *In situ* visualization of proteins of interest at single cell level is attractive in cell biology, molecular biology and biomedicine, which usually involves photon, electron or X-ray based imaging methods. Herein, we report an optics-free strategy that images a specific protein in single cells by time of flight-secondary ion mass spectrometry (ToF-SIMS) following genetic incorporation of fluorine-containing unnatural amino acids as a chemical tag into the protein via genetic code expansion technique. The method was developed and validated by imaging GFP in E. coli and human HeLa cancer cells, and then utilized to visualize the distribution of chemotaxis protein CheA in E. coli cells and the interaction between high mobility group box 1 protein and cisplatin damaged DNA in HeLa cells. The present work highlights the power of ToF-SIMS imaging combined with genetically encoded chemical tags for *in situ* visualization of proteins of interest as well as the interactions between proteins and drugs or drug damaged DNA in single cells.

A variety of chemical and biological reactions occurred inside cells at every moment. It is crucial, yet hard to monitor in situ the variations, for example the turnover and/or translocation of proteins of interest at subcellular levels.[1] Classic methods to visualize proteins in single cells mainly rely on photon, electron or X-ray based techniques, for example, confocal fluorescence imaging, electron microscopy and X-ray microscopy. During the past decades, mass spectrometry imaging (MSI) with various ion sources has attracted widespread attention for visualization the proteins of interest in tissues and in cells.[2] Among these MSI techniques, secondary

ion mass spectrometry (SIMS), including NanoSIMS and ToF-SIMS, with nanoscale spatial resolution, has gained increasing application in cell biology, molecular biology and biomedicine research fields, e.g., studies on subcellular distribution and turnover of proteins, in combination with stable isotopic labeling technique.[3] Until recently, a series of methods using isotopic labeling,[4] fluorine-tagged immuno-staining,[5] and click reaction labeling[6] have been reported to successfully visualize specific proteins at single cell level by SIMS imaging.

In 2001, Schultz and co-workers developed a genetic code expansion (GCE) technique, making it possible to site-specifically incorporate unnatural amino acids (UAAs) into a protein of interest in either prokaryotic or eukaryotic living cells through an orthogonal transfer RNA (tRNA) and aminoacylated RNA synthetase (aaRS) system, i.e. tRNA/aaRS system.[7] The modified proteins with UAAs enable various studies on proteins, such as protein site-directed modification, protein turnover, protein structure and function, etc.[8] Recently, GCE technique has also been applied to introduce indirectly fluorine-/boron-containing chemical tags via intracellular click reaction of F-/B-containing probes with UAA propargyl-L-lysine (PRK) inserted to proteins. This allowed precise visualization of the proteins of interest by NanoSIMS.[6, 9] To avoid the complex intercellular click reaction, in this work, we attempt to encode directly F-containing UAAs as chemical tags into specific proteins by GCE, for in situ visualization of proteins by ToF-SIMS imaging in cells.

Firstly, we chose green fluorescent protein (GFP) as a model protein to incorporate F-containing UAA into a protein expressed in E. coli cells for ToF-SIMS imaging. The plasmid, which can express GFP cooperated with UAA, was constructed and optimized in-house. Briefly, a stop codon (TAG) was inserted in the front of the GFP sequence in the plasmid, and the TAG-GFP sequence was cut and fused into pET-28a plasmid. The resulting plasmid was mixed with the pSupAR-Mb-DiZPK-RS plasmid, and transfected into BL21 cells to construct strain A, which were subsequently induced to express GFP containing the given UAAs in the presence of the exogenous orthogonal tRNA/aaRS system. More details for construction of various plasmids are given in the Supporting Information.

In order to verify whether GFP can be labeled in E. coli cells transfected with dual plasmids, the strain A was induced to express proteins in the presence of either N6-carbobenzyloxy-lysine (N6-CBZ-K) or 3-Trifluoromethyl-phenylalanine (Tf-phe) (Figure 1A). For comparison, the strain A was also induced for protein expression in the absence of UAAs. Meanwhile the strain B, which was transfected with the wild type GFP plasmid, was also induced to express the wild type GFP protein under the conventional conditions. All three types of GFP were characterized by mass spectrometry and SDS-PAGE (Figure 1B and Figure S1 in the Supporting Information), confirming the successful expression of the expected proteins. Laser scanning confocal microscopy (LSCM) images of cells further demonstrated that the UAA modified GFP was successfully expressed in E. coli in the presence of UAAs, since only the modified GFP expressed in full-length can give fluorescence, whereas little GFP was generated in the absence of UAAs (Figure S2). Moreover, the

fluorescence images showed that the UAA-modified GFP protein evenly distributed throughout the cells as the wild type GFP did.

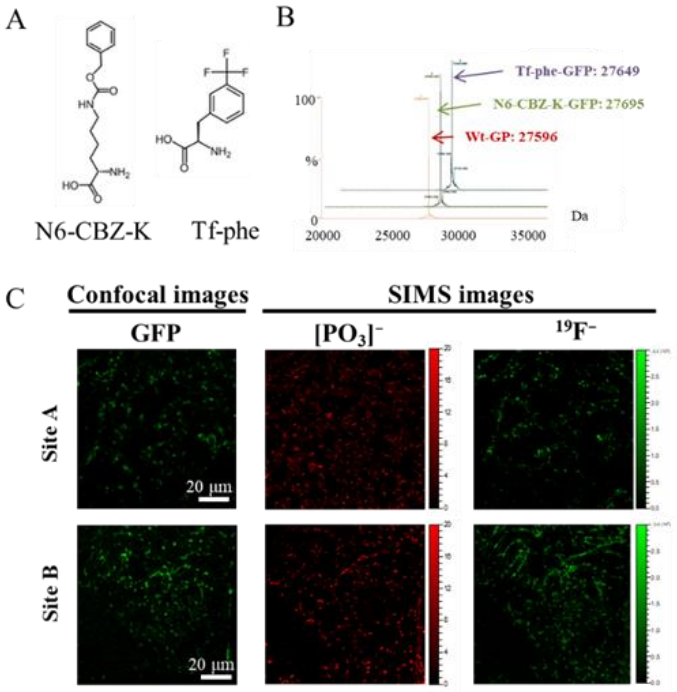

Figure 1. (A) Chemical structure of unnatural amino acids (UAAs) N6-carbobenzyloxy-lysine (N6-CBZ-K) and trifluoromethyl-phenylalanine (Tf-phe). (B) Mass spectra of wild type GFP (expected molecular weight (MWexp)/observed molecular weight (MWobs): 27596/27596), N6-CBZ-K modified GFP (MWexp/MWobs: 27695/27695) and Tf-phe modified GFP (MWexp/MWobs: 27647/27649). (C) LSCM and ToF-SIMS images of strain A expressing GFP in the presence of Tf-phe as a UAA.

Next, the strain A was successively imaged by LSCM and ToF-SIMS (Figure 1C) using a protocol developed previously by us.[10] For ToF-SIMS imaging of E. coli cells, the signal of $[PO_3]^-$ ions was collected to render the morphology of the cells, while the signal of $19F^-$ ions was collected to localize Tf-phe incorporated GFP. The results indicated that the strain A significantly expressed Tf-phe modified GFP as evidenced by the strong signal of $19F^-$ ions (Figure 1C). Moreover, the SIMS signal of $19F^-$ ions right match the fluorescence signal of GFP, further verifying the incorporation of F-containing UAA Tf-phe, i.e. the F-chemical tag, to GFP. The control experiments demonstrated that the signal of $19F^-$ ions was not detected by ToF-SIMS when the strain A was induced to express GFP in the presence of N6-CBZ-K or in the absence of UAAs (Figure S3). When the strain A was cultured in the presence of Tf-phe but without induction, little $19F^-$ signal was detected, indicating that no Tf-phe residues were inserted to GFP protein, and that the free UAA did not interfere the detection of genetically encoded chemical tags as the free UAA was washed out from the cells (Figure S3). These results proved that the proposed strategy is feasible, and the chosen genetic encoded chemical tags can be effectively

incorporated into GFP for visualization of the protein in E. coli cells by ToF-SIMS imaging.

Following the success in mapping GFP in E. coli cells, we proceeded to utilize the developed method to locate a protein of interest, cytoplasmic protein CheA in E. coli. The bacterial chemotaxis protein CheA was reported to localize at the poles of E. coli cells.[11] The E. coli strains were outfitted with a dual-plasmid expression system, which is consisted of one plasmid (pSupAR-Mb-DiZPK-RS) directly expressed the RS and the corresponding tRNA and a second coding for inducible expression of the protein. Using this expanded genetic code system, CheA protein was expressed with F-chemical tags in the same fashion as was GFP (vide supra), as evidenced by gel electrophoresis and mass spectrometry analysis (Figure S4). To find out whether CheA localizes at the poles of E. coli cells as reported previously,[11c, 11d] we mapped the Tf-phe modified CheA by ToF-SIMS. Similarly, $[PO_3]^-$ ions was imaged to profile the shape of cells. As shown in Figure 2, the ToF-SIMS images of $19F^-$ ions demonstrated that CheA indeed present polar localization. This result again proved that ToF-SIMS imaging coupled with genetically encoded F-chemical tag is an efficient tool to visualize the proteins of interest in prokaryotes.

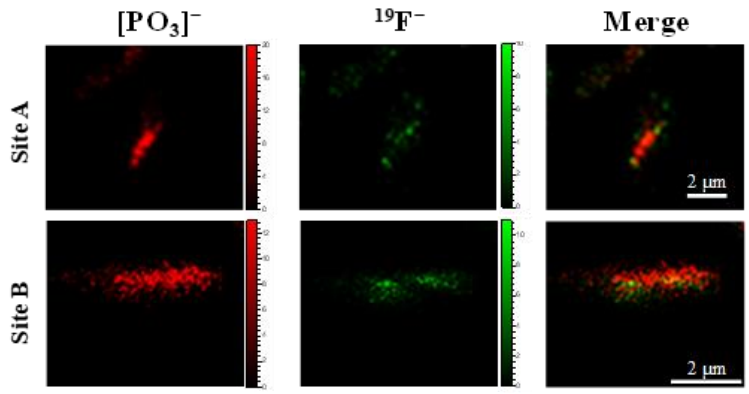

Figure 2. ToF-SIMS images of CheA protein expressed in E. coli in the presence of Tf-phe.

Next, we expanded the genetical encoded chemical tag strategy to label specific proteins in eukaryotic cells for the localization of the proteins by ToF-SIMS. Again, GFP was selected as a model protein. The plasmid which expresses GFP was mutated at site 151 to generate a TAG and ligated into a plasmid vector which can be induced to express the protein in eukaryotic cells. The resulted pCMV-151TAG-GFP plasmid was transfected into eukaryotic cells and induced to express GFP.

Three kinds of eukaryotic cells, human cancer cells HeLa, A549 and MCF-7 were transfected by plasmids pCMV-151TAG-GFP and pCMV-Mb-PylRS-WT, and cultured in the absence of UAAs or in the presence of N6-CBZ-K or Tf-phe, respectively. As shown in Figure S5, all three cell lines expressed GFP in the presence of UAAs, but at a lower efficiency (Figure S2). In order to promote the labeling efficiency of GFP by the F-chemical tag, pCMV-Mb-PylRS-WT plasmid was mutated at site 313 for the higher incorporation rate of Tf-phe into GFP.[12] The confocal

fluorescence imaging on living HeLa cells demonstrated that pCMV-Mb-PylRS-313Mut enabled the HeLa cells to express Tf-phe-modified GFP at a higher efficiency than N6-CBZ-K-modified GFP (Figure S6), indicating that pCMV-Mb-PylRS-313Mut plasmid favored incorporation of Tf-phe over N6-CBZ-K into proteins in eukaryotic cells.

Next, HeLa cells were transfected with plasmids pCMV-TAG-GFP and pCMV-Mb-PylRS-313Mut, and cultured on a house-made addressable silicon wafer in our lab[10] in the presence of Tf-phe as UAA. After fixed on the silicon wafer and stained by the nucleus dye Hoechst33342, correlated LSCM and ToF-SIMS imaging was performed on the lyophilized HeLa cells. As shown in Figure 3, the F-tagged GFP was effectively expressed and distributed in the entire cells. These results proved that the genetically encoded F-chemical tags can be applied to label proteins of interest in both living cells and fixed human cells for ToF-SIMS imaging.

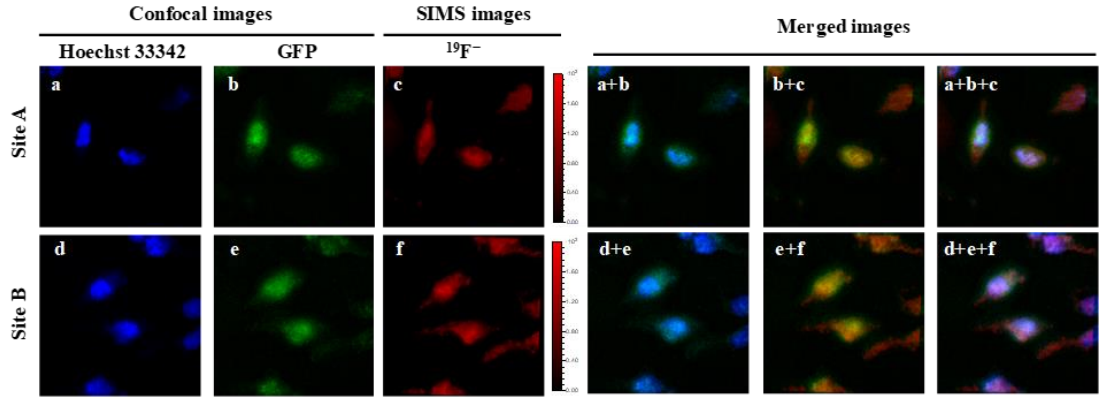

Figure 3. Correlated LSCM and ToF-SIMS images of F-tagged GFP expressed in HeLa cells in the presence of Tf-phe as UAA.

In order to further validate the application of the developed F-chemical tags for ToF-SIMS imaging of proteins in human cells, more control experiments were carried out. When no UAA was added or no plasmids was transfected during cell culturing, the HeLa cells did not express the F-tagged GFP, neither fluorescence signal of GFP nor SIMS signal of $19F^-$ was detected (Figures S7). These results further illustrated the feasibility of the genetically encoded F-chemical tags for ToF-SIMS imaging of proteins in eukaryotes cells.

In situ study of the recognition and interaction between proteins and drugs or proteins and drug-damaged DNA in cells has always been the biggest concern for drug discovery.[13] High mobility group box 1 (HMGB1) is a non-sequence-specific DNA-binding protein involved in the regulation of many functions of the nucleus, including transcription, replication, recombination, and general chromosomal remodeling.[14] In vitro researches showed that the specific binding of HMGB1 to the damage site of DNA by DNA-targeting anticancer drug cisplatin prevented the 1,2-crosslinked DNA lesion from being recognized and repaired by Nucleotide Excision Repair (NER) system. This, as a consequence, induces cell apoptosis and

kills cancer cells.[13b, 14b, 15] However, this unique recognition and interaction has not been proved in single cell level. Thus, we applied the ToF-SIMS imaging method described above to visualize in situ and verify the interaction between HMGB1 and 1,2-cisplatin-crosslinked DNA in cells. To this end, we doubly labeled HMGB1 via genetically encoded green fluorescence protein and F-chemical tag Tf-phe. To avoid potential effect of UAA-insertion on the recognition we concern, the UAA Tf-phe was genetically encoded at site 182 in HMGB1. The details of genetic engineering procedure are given in the Supporting Information.

HeLa cells expressing Tf-phe modified GFP-HMGB1 were then incubated with 50 μM cisplatin for 24 h, and imaged by LSCM first. The fluorescence signal was observed in the nuclear area (Figure S8) instead of the entire cells as the solo GFP (Figure 3), indicating the successful expression of the GFP-HMGB1 fused protein. Next, HeLa cells were transfected and incubated with 50 μM or 100 μM cisplatin for 24 h, and then lyophilized, directly followed by ToF-SIMS imaging. Both the magnified single-cell image or a bit crowd of single cells image demonstrated that cisplatin mainly distributed in the nuclear areas as rendered by the signal of $[PtCN]^-$ ions,[16] so did HMGB1 as shown by the signal of $19F^-$ ions (Figure 4). More importantly, the fused images of $[PtCN]^-$ and $19F^-$ ions show significant co-localization of cisplatin and HMGB1 in nuclei, indicating the formation of HMGB1-cisplatin-DNA ternary complex inside cells. This is in consistence with our recent report that the binding of HMGB1 to the cisplatin damaged DNA could be observed by successive application of laser scanning confocal fluorescence microscopy and ToF-SIMS in single cells.[17]

The Phe37 residue of HMGB1 has been reported to dominate the recognition between HMGB1 and cisplatin damaged DNA.[14b] Thus, the mutation at Phe37 is expected to disrupt this unique recognition. To further verify the reliability of our ToF-SIMS imaging described above, we carried out a control experiment with a site-specific mutation at phe37 of HMGB1. Similarly, the mutant HMGB1(F37A) protein was expressed as a GFP-fused protein, and the F-containing UAA Tf-phe was genetically encoded at site 182 of HMGB1. The HeLa cells expressing Tf-phe modified GFP-HMGB1(F37A) were incubated with 50 μM cisplatin for 24 h, and imaged by LSCM, the results (Figure S9) showed that although the fused HMGB1 protein is still mainly present in the nucleus, it also has a certain distribution in the cytoplasm. Figure S10 further showed the results of transfected cells imaged directly by ToF-SIMS after lyophilizing. Although cisplatin rendered by the signal of $[PtCN]^-$ ions appear to distribute homogeneously in the nuclei as happened in the HeLa cells expressing wild-type GFP-HMGB1 (Figure 4), little co-localization of cisplatin with the GFP-HMGB1(F37A) protein indicated by the signal of $19F^-$ ions was observed. These results not only proved that the mutation at Phe37 of HMGB1 indeed disables its specific affinity binding to cisplatin damaged DNA, but also confirm that the genetically encoded chemical tag enables ToF-SIMS to be a powerful tool for visualizing proteins of interest in single cells.

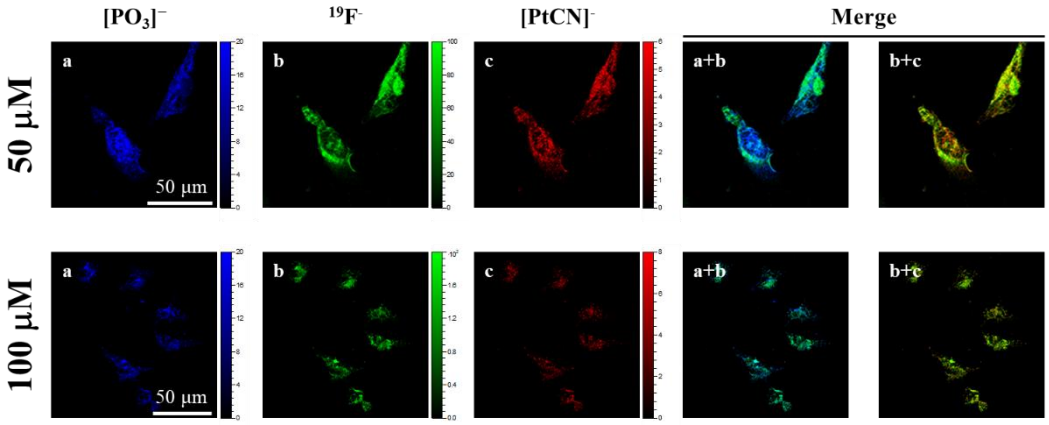

Figure 4. ToF-SIMS images of HeLa cells which expressed Tf-phe modified GFP-HMGB1 and were incubated with different concentrations of cisplatin for 24 h.

In summary, we developed an optics-free strategy that chemically tagged proteins of interest by F-containing UAA via the genetic code expansion technique, and then *in situ* visualize the modified proteins in single cells by ToF-SIMS imaging. This proposed method was developed and validated by imaging trifluoromethyl-phenylalanine (Tf-phe) modified GFP in both prokaryotic and eukaryotic cells, and applied to visualize the chemotaxis protein CheA in E. coli and HMGB1 protein in human HeLa cells. The results unambiguously demonstrated that CheA localized at the poles of E. coli cells. More importantly, as SIMS imaging has numerous mass channels for simultaneous imaging of multiple targets, so the interaction of a specific protein with non-fluorescent small or macromolecules can thus be observed. In this work, by combining the imaging of metallodrug cisplatin and Tf-phe modified GFP-HMGB1 in the same single cells by ToF-SIMS, we in situ visualized for the first time the recognition and interaction between HMGB1 and cisplatin damaged DNA at single cell level. This work demonstrates the feasibility and advantages of ToF-SIMS imaging coupled to genetically encoded chemical tags for in situ investigation of proteins of interest in cells. We anticipate this novel strategy could be widely applied to visualize proteins as well as the reactions of proteins with drugs or drug damaged DNA in cells.

ASSOCIATED CONTENT

Supporting information

General considerations, supplementary methods, and Figures.

AUTHOR INFORMATION

Corresponding Author

* Email: fuyi.wang@iccas.ac.cn; yaozhao@iccas.ac.cn

Notes

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Prof. Dr. Peng R. Chen (College of Chemistry and Molecular Engineering, Peking University) for kindly gifting pSupAR-Mb-DiZPK-RS plasmid and pCMV-Mb-PylRS-WT plasmid. We also acknowledge the technical support from Dr. Yu Lin (Institute of Chemistry, Chinese Academy of Sciences) for the plasmid optimization. We thank the NSFC (grant nos. 21927804, 21575145, 91543101, 21904128, 21635008, 21790392, 21790390) and the National Key R&D Program of China (grant no. 2018YFA0800903) for support. Y.Z. thank the Youth Innovation Promotion Association of Chinese Academy of Sciences (grant no. 2017051).

REFERENCES

[1]a) E. Lundberg, G. H. H. Borner, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 285-302; b) L. Xue, I. A. Karpenko, J. Hiblot, K. Johnsson, *Nat. Chem. Bio.* **2015**, *11*, 917-923; cR. Aebersold, M. Mann, *Nature* **2016**, *537*, 347-355.

[2] a) P. D. Piehowski, Y. Zhu, L. M. Bramer, K. G. Stratton, R. Zhao, D. J. Orton, R. J. Moore, J. Yuan, H. D. Mitchell, Y. Gao, B.-J. M. Webb-Robertson, S. K. Dey, R. T. Kelly, K. E. Burnum-Johnson, *Nat. Commun.* **2020**, *11*, 8; b) M. Kompauer, S. Heiles, B. Spengler, *Nat. Met.* **2017**, *14*, 90-96; c) A. Bodzon-Kulakowska, P. Suder, *Mass Spectrom. Rev.* **2016**, *35*, 147-169; d) L. Keren, M. Bosse, D. Marquez, R. Angoshtari, S. Jain, S. Varma, S.-R. Yang, A. Kurian, D. Van Valen, R. West, S. C. Bendall, M. Angelo, *Cell* **2018**, *174*, 1373-1387.e1319; e) J. Han, H. Permentier, R. Bischoff, G. Groothuis, A. Casini, P. Horvatovich, *TrAC, Trends Anal. Chem.* **2019**, *112*, 13-28; f) E. K. Neumann, T. D. Do, T. J. Comi, J. V. Sweedler, *Angew. Chem. Int. Ed.* **2019**, *58*, 9348-9364; g) N. T. N. Phan, X. Li, A. G. Ewing, *Nat. Rev. Chem.* **2017**, *1*, 0048.

[3] L. Zhang, A. Vertes, *Angew. Chem. Int. Ed.* **2018**, *57*, 4466-4477.

[4] L. Carlred, A. Gunnarsson, S. Solé-Domènech, B. Johansson, V. Vukojević, L. Terenius, A. Codita, B. Winblad, M. Schalling, F. Höök, P. Sjövall, *J. Am. Chem. Soc.* **2014**, *136*, 9973-9981.

[5] S. Kabatas, P. Agüi-Gonzalez, R. Hinrichs, S. Jähne, F. Opazo, U. Diederichsen, S. O. Rizzoli, N. T. N. Phan, *J. Anal. At. Spectrom.* **2019**, *34*, 1083-1087.

[6] a) I. C. Vreja, S. Kabatas, S. K. Saka, K. Krohnert, C. Hoschen, F. Opazo, U. Diederichsen, S. O. Rizzoli, *Angew. Chem. Int. Ed.* **2015**, *54*, 5784-5788; b) S. Kabatas, P. Agui-Gonzalez, K. A. Saal, S. Jahne, F. Opazo, S. O. Rizzoli, N. T. N. Phan, *Angew. Chem. Int. Ed.* **2019**, *58*, 3438-3443.

[7] a) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498-500; b) L. Wang, P. G. Schultz, *Angew. Chem. Int. Ed.* **2004**, *44*, 34-66.

[8] a) X. Xie, X. M. Li, F. Qin, J. Lin, G. Zhang, J. Zhao, X. Bao, R. Zhu, H. Song, X. D. Li, P. R. Chen, *J. Am. Chem. Soc.* **2017**, *139*, 6522-6525; b) J. Wang, S. Zheng, Y. Liu, Z. Zhang, Z. Lin, J. Li, G. Zhang, X. Wang, J. Li, P. R. Chen, *J. Am. Chem. Soc.* **2016**, *138*, 15118-15121; c) I. Drienovská, G. Roelfes, *Nat. Catal.* **2020**, DOI:10.1038/s41929-41019-40410-41928; d) J. W. Chin, *Nature* **2017**, *550*, 53-60; e) Y. Zheng, T. L. Lewis, P. Igo, F. Polleux, A. Chatterjee, *ACS Synth. Bio.* **2017**, *6*, 13-18; f) F. Zhang, Q. Zhou, G. Yang, L. An, F. Li, J. Wang,

Chem. Commun. **2018**, *54*, 3879-3882.

[9] S. Kabatas, I. C. Vreja, S. K. Saka, C. Hoschen, K. Krohnert, F. Opazo, S. O. Rizzoli, U. Diederichsen, *Chem. Commun.* **2015**, *51*, 13221-13224.

[10] S. Liu, W. Zheng, K. Wu, Y. Lin, F. Jia, Y. Zhang, Z. Wang, Q. Luo, Y. Zhao, F. Wang, *Chem. Commun*. **2017**, *53*, 4136-4139.

[11] a) X. Ding, Q. He, F. Shen, F. W. Dahlquist, X. Wang, *J. Bacteriol*. **2018**, *200*, e00052-00018; b) X. Wang, C. Wu, A. Vu, J.-E. Shea, F. W. Dahlquist, *J. Am. Chem. Soc.* **2012**, *134*, 16107-16110; c) S. H. Ho, D. A. Tirrell, *J. Am. Chem. Soc*. **2016**, *138*, 15098-15101; d) Maddock, L. Shapiro, *Science* **1993**, *259*, 1717.

[12] A. Hohl, R. Karan, A. Akal, D. Renn, X. Liu, S. Ghorpade, M. Groll, M. Rueping, J. Eppinger, *Sci. Rep*. **2019**, *9*, 11971.

[13] a) Y. Jung, S. J. Lippard, *Chem. Rev.* **2007**, *107*, 1387-1407; b) D. Wang, S. J. Lippard, *Nat. Rev. Drug Discov.* **2005**, *4*, 307-320.

[14] a) D. L. Andreas D. Baxevanis, *Nucleic Acids Res.* **1995**, *23*, 1604-1613; b) U.-M. Ohndorf, M. A. Rould, Q. He, C. O. Pabo, S. J. Lippard, *Nature* **1999**, *399*, 708-712; c) Z. Du, Q. Luo, L. Yang, T. Bing, X. Li, W. Guo, K. Wu, Y. Zhao, S. Xiong, D. Shangguan, F. Wang, *J. Am. Chem. Soc*. **2014**, *136*, 2948-2951.

[15] C. X. Zhang, P. V. Chang, S. J. Lippard, *J. Am. Chem. Soc.* **2004**, *126*, 6536-6537.

[16] Y. Zhang, Q. Luo, W. Zheng, Z. Wang, Y. Lin, E. Zhang, S. Lu, J. Xiang, Y. Zhao, F. Wang, *Inorg. Chem. Front.* **2018**, *5*, 413-424.

[17] Y. Lin, K. Wu, F. Jia, L. Chen, Z. Wang, Y. Zhang, Q. Luo, S. Liu, L. Qi, N. Li, X. Fang, P. Dong, F. Gao, Y. Zhao, F. Wang, **2020**, https://doi.org/10.26434/chemrxiv.11662335.v3.