Cysteine-specific Chemical Proteomics: From Target Identification to Drug Discovery

Dominic G. Hoch, Daniel Abegg, Chao Wang, Anton Shuster, and Alexander Adibekian*

Abstract: Our laboratory focuses on chemical proteomics-enabled discovery of new cysteine-reactive small molecules with intriguing biomedical activities as well as identification and detailed characterization of their proteomic targets. In this overview article, we summarize our progress since 2013 in this research field. We have developed a novel mass spectrometry-based chemoproteomic method that allows detection and monitoring of up to ~3000 reactive cysteines in any cellular proteome. This is achieved *via* strategic use of two clickable, cysteine-reactive chemical probes with complementary substrate selectivity profiles, iodoacetamide and ethynyl benziodoxolone. Using this method, we have been able to identify the direct biological targets of curcumin, a diarylheptanoid natural product with anticancer activity, and deoxyelephantopin, a highly cytotoxic natural sesquiterpene lactone. Furthermore, we have developed chloromethyl triazoles (CMTs) as a novel chemical scaffold for cysteine-reactive inhibitors that can be accessed from commercially available substrates in only two chemical steps. From a small collection of chloromethyl triazoles, we have identified compound AA-CW236 as the first non-pseudosubstrate inhibitor of MGMT, a DNA repair protein that renders several devastating cancer forms resistant to chemotherapy.

Keywords: Chemical proteomics · Cysteine-reactive small molecules · Mass spectrometry · Natural products · Proteomic profiling

Alexander Adibekian obtained his BSc and MSc from the University of Hannover, Germany. In 2008, he completed his PhD studies with Prof. Peter H. Seeberger at the ETH Zürich. Alex subsequently moved to San Diego to work as a postdoctoral fellow with Prof. Benjamin Cravatt at The Scripps Research Institute. In 2012, he started his independent career in the Department of Biochemistry at the University of Geneva. In August 2013, Alex joined the Department of Organic Chemistry at the same institution and the National Centre of Competence in Research (NCCR) in Chemical Biology as Assistant Professor.

Correspondence: Prof. Dr. A. Adibekian
Department of Organic Chemistry
NCCR Chemical Biology
University of Geneva
30 quai Ernest-Ansermet, Geneva
E-mail: alexander.adibekian@unige.ch

Cysteines are indispensable for the correct functioning of proteins in our cells. The chemical reactivity of each single cysteine in the proteome is fine-tuned to perform a specific biochemical task. So-called hyperreactive cysteines are often involved in nucleophilic and reductive catalysis and are found in diverse families of enzymes of biomedical importance such as proteases, kinases and oxidoreductases.[1] Hyperreactive cysteines can be chemoselectively targeted with moderately reactive electrophilic inhibitors such as chloro-acetamides and α,β-unsaturated amides. However, the fact that such drugs may also react with proteins other than the desired target, has caused concerns regarding the long-term health risk of using covalent drugs. Accordingly, many high-throughput drug discovery endeavors have historically tended to avoid irreversible inhibitors. Nonetheless, covalent binders often display certain advantageous properties in comparison to their reversible counterparts, namely easily achievable high potency and a strongly prolonged duration of action.[2]

For example, recent studies suggest that there is significantly less risk to develop resistance against covalent inhibitors, which is a major advantage for application in areas such as infectious diseases and oncology. A number of covalent cysteine-reactive inhibitors have recently been approved by the FDA and this showcases the potential of this relatively novel category of drugs.

Since 2013, our laboratory has focused on chemical proteomics-driven discovery of new cysteine-reactive small molecules with intriguing biomedical activities as well as the detailed characterization of their proteomic targets. Particularly the identification of targets of bioactive molecules is not an easy task and often requires synergy of methods from different scientific disciplines ranging from organic synthesis to cell biology and mass spectrometry-based proteomics.[3] A relatively new and currently one of the most popular techniques for target identification is the so-called competitive proteomic profiling method[4] that relies on proteome-wide labeling of reactive proteomic sites such as serines, cysteines or lysines with alkyne- or azide-tagged electrophilic probes (Fig. 1). The probe-labeled proteome is then reacted with an azide- or alkyne-derivatized fluorescent dye through copper(ɪ)-catalyzed alkyne-azide cycloaddition (CuAAC), separated by SDS-PAGE and visualized by in-gel fluorescence scanning. If the proteome is first incubated with a covalently binding inhibitor of interest, the inhibitor will react with its natural proteomic targets and the inhibitor-blocked proteomic sites will not be labeled by the electrophilic probe thus resulting in fluorescent signal loss. Alternatively, probe-labeled proteins can also be 'click'-tagged with biotin, en-

riched with streptavidin beads, digested and directly identified *via* LC-MS/MS, thus enabling target identification in high-throughput fashion. Iodoacetamide (IAA) is widely seen as gold standard among cysteine-reactive chemical probes,[5] but this probe is relatively unstable in aqueous buffers and is only moderately reactive, resulting in labeling of only a small fraction of functional cysteines. Furthermore, iodoacetamide also tends to crossreact with proteomic lysines mimicking natural lysine ubiquitination sites in subsequent MS analysis. Therefore, there is currently a strong demand for new cysteine-reactive probes that show complementary selectivity to IAA to allow much more comprehensive proteomic profiling of small molecule targets.

A Novel Cysteine-specific Proteomic Profiling Method

It was known that ethynyl benziodoxolones (EBX reagents), developed by Jérôme Waser and coworkers at the EPFL, readily alkynylate thiols and that this reaction proceeds selectively in the presence of many other nucleophilic functional groups.[6] However, these transformations usually require water/organic solvent mixtures and stoichiometric amounts of organic base. Therefore, we sought to investigate whether EBX reagents can also serve as useful bioorthogonal probes for proteomic profiling of cysteines. We first performed a series of control experiments that showed that the azide-functionalized EBX reagent JW-RF-010 is not only water-stable but is also five times more reactive towards cysteine-containing proteins than iodoacetamide (Fig. 2). Moreover, we were able to demonstrate that JW-RF-010 can be used

to label proteins directly in live cells.[7] To identify the JW-RF-010- but also IAA-labeled proteomic sites with high precision and sensitivity, we established a capture-and-release strategy that allowed selective enrichment of probe-labeled proteins, on-bead digestion and release of probe-bound peptides from the beads by simple UV-irradiation. This strategy allowed identification of thousands of probe-labeled proteomic sites and also the assessment of the chemoselectivity of both probes. Our data indicate that JW-RF-010 is highly selective towards cysteines (97%), while only 3% of alkynyl adducts were found on other nucleophilic amino acids such as aspartic acid, glutamic acid, threonine and serine. Moreover, 38% of JW-RF-010-enriched peptides were unique and were not found in the IAA dataset. Both probes show great complementarity in terms of substrate selectivity and should ideally be used in combination to allow best proteomic coverage.

To demonstrate the advantage of the dual probe approach, we decided to investigate the biological targets of curcumin, a diarylheptanoid natural product with anticancer and anti-inflammatory properties isolated from turmeric (*Curcumin longa*) (Fig. 3).[8] Overall, we identified 42 direct targets of curcumin at 30 µM concentration in HeLa cells. Among these targets, 19 proteins were only enriched by IAA alkyne and 16 targets were exclusively labeled by JW-RF-010. This result clearly demonstrates the importance of using both probes in parallel for cysteine profiling. One of the identified targets, casein kinase I gamma (CSNK1G), could at least partially explain the reported anticancer activity of curcumin. It is known that siRNA knockdown of CSNK1G3 reduces the activity-enhancing phosphorylation on Ser473 of Akt, a serine-threonine kinase that acts as

a suppressor of apoptosis and represents a major anticancer target.[9] Curcumin modifies the Cys51 of CSNK1G3 that is located in the so-called P-loop in close proximity to the ATP-binding pocket of this kinase (Fig. 3). Additional biochemical experiments confirmed that curcumin inhibits the ATP hydrolysis by CSNK1G3.

Identification of the Biological Targets of Deoxyelephantopin

In general, cysteine-reactive natural products, although often too toxic for clinical use, find application in life sciences as inhibitors of important cellular pathways. For example, helenalin and parthenolide block the NF-κB pathway by covalently engaging this transcription factor.[10] Thapsigargin is a covalent inhibitor of SERCA (sarco/endoplasmic reticulum Ca^{2+} ATPase) and is broadly used in cell biology to increase cytosolic calcium concentrations.[11] Extracts of the plant *Elephantopus scaber* have been used in traditional asian medicine for long time and the sesquiterpene lactone deoxyelephantopin (DEP) represents one of the bioactive components of this plant extract.[12] For instance, deoxyelephantopin is as effective as paclitaxel in a mouse breast cancer model.[13] Moreover, surface plasmon resonance studies detected interaction between deoxyelephantopin (DEP) and PPARγ (peroxisome proliferator-activated receptor gamma),[14] a transcription factor that is heavily involved in the pathogenesis of obesity, diabetes and atherosclerosis. However, because these studies were performed *in vitro* with purified PPARγ, it was not clear whether this finding has any physiological relevance. The chemical structure of DEP contains three electrophil-
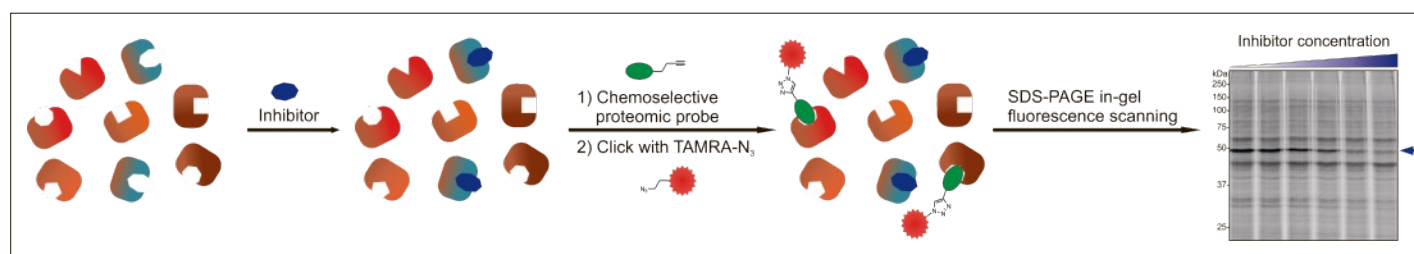

Fig. 1. General workflow of a competitive proteomic profiling experiment to identify inhibitor targets.
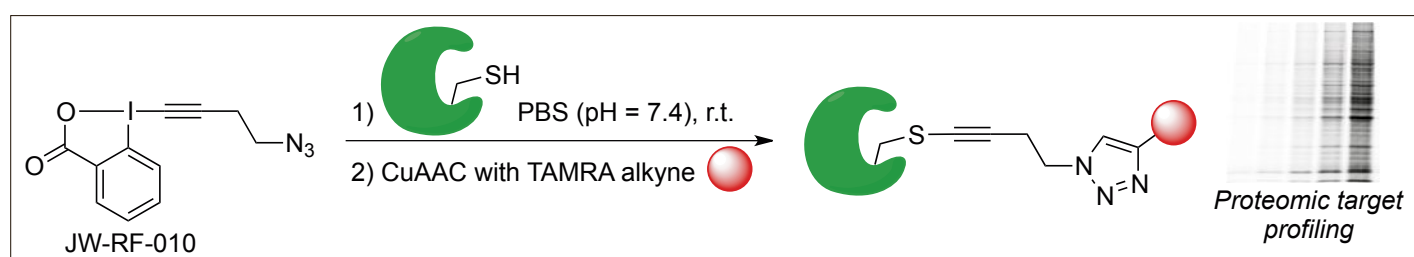

Fig. 2. Alkynylation of cysteine-containing proteins with the clickable EBX probe JW-RF-010 for proteomic profiling experiments.
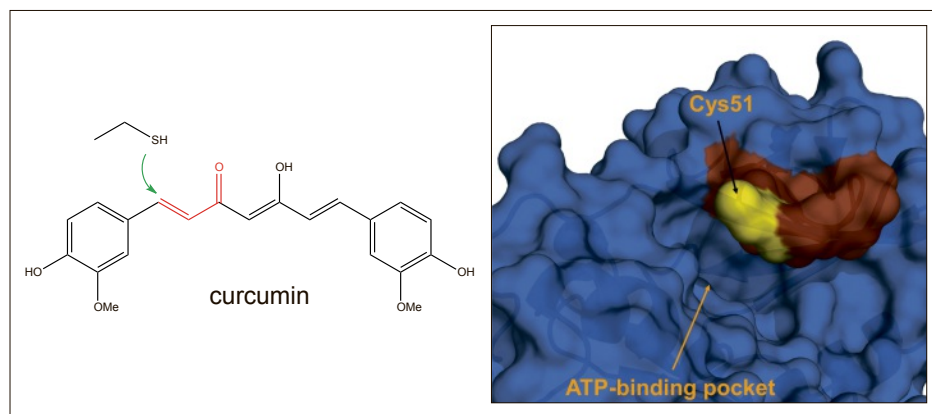
Fig. 3. Left: chemical structure of curcumin. Right: protein structure of H. sapiens CSNK1G3 (PDB: 2IZS) showing the ATP-binding pocket of the kinase, the P-loop (brown) and the Cys51 (yellow). The image was created using PyMOL (V1.7.2.1).

ic sites that may potentially react with proteomic cysteines. Driven by the curiosity to know whether the bioactivity of DEP could indeed be explained by its chemical reactivity with thiols, we performed proteome-wide identification of cysteine-specific targets (TargetID) of deoxyelephantopin and some unnatural DEP analogs that were synthesized by the Winssinger laboratory at the UniGE (Fig. 4). We indeed identified the proteins cystathionine beta-synthase (CBS), src substrate cortactin (CTTN) and cystatin-B (CSTB) as cancer-relevant direct targets of DEP in MCF7 cells (breast cancer).[15] More importantly, by repeating the chemoproteomic analysis in caco-2 cells (colon cancer), we confirmed that DEP is also capable of covalently binding PPARγ under physiological conditions. We confirmed this result by overexpressing PPARγ and successfully performing gel-based labeling of this protein with a clickable deoxyelephantopin derivative. Moreover, pretreatment of PPARγ with iodoacetamide completely abolished the labeling, thus indicating that DEP indeed interacts with one of the cysteine residues of PPARγ. To better understand the effect of DEP on PPARγ activity in cells, we performed a global proteomics experiment to identify dysregulated proteins after 24 h treatment with deoxyelephantopin, PPARγ agonist rosiglitazone or PPARγ antagonist T0070907. We identified 25 upregulated and 18 downregulated proteins that were common between DEP and T0070907
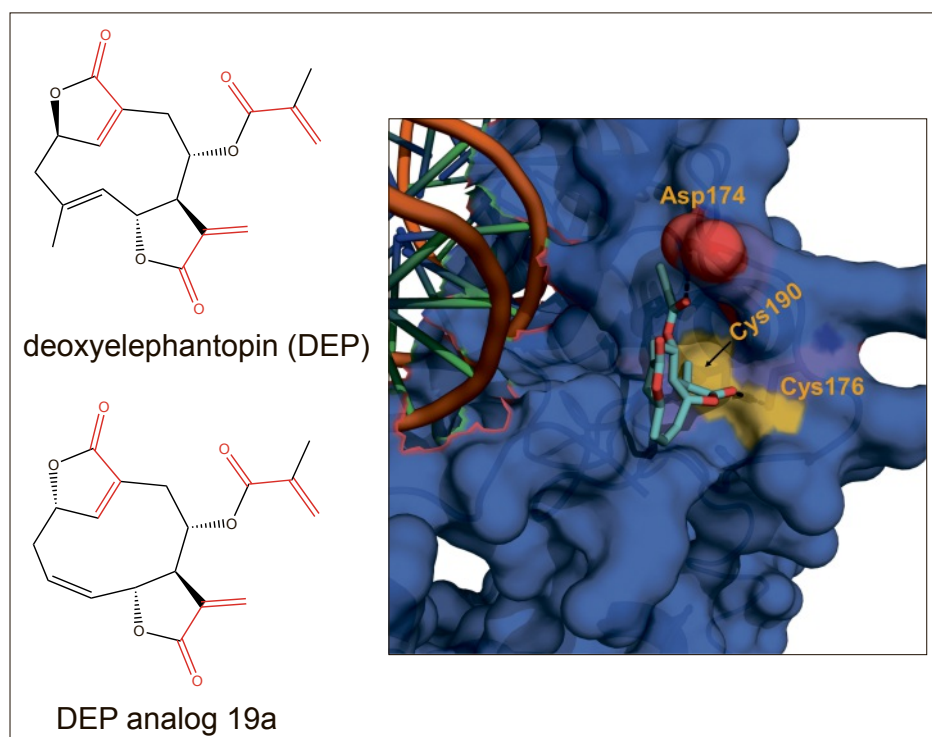
datasets, thus strongly indicating that deoxyelephantopin acts in cells as PPARγ antagonist. Importantly, DEP treatment led to strong downregulation of retinol-binding protein 4 (RBP4), an adipokine and a major therapeutic target that contributes to insulin resistance in obesity and type 2 diabetes.[16] This result suggests that pharmacologically improved DEP analogs may indeed find their way into clinics as insulin resistance reducing drugs for applications in various metabolic disorders. Finally, we performed competitive proteomics experiments to determine the exact cysteine site on PPARγ that interacts with DEP and analogs. Detailed interrogation of MS/MS spectra resulted in the identification of Cys190 as the putative site of covalent modification. The X-ray structure of PPARγ shows that Cys190 is coordinated to a Zn^{2+} ion in a zinc finger motif as part of the DNA-binding domain of this transcription factor (Fig. 4). We hypothesize that binding of deoxyelephantopin and analogs to Cys190 may potentially disrupt the interaction between PPARγ and DNA and thus explain the observed antagonizing effect.

Chloromethyl Triazoles: A Novel Cysteine-reactive Chemical Scaffold

While the TargetID approach with natural products often yields surprising and exciting results, natural products rarely show satisfactory inhibitory potency and target selectivity to qualify for deeper pharmacological investigations. This usually can only be achieved with fully synthetic small molecules that have undergone several rounds of optimization. We are developing novel chemical strategies that allow rapid access to collections of fully synthetic and structurally diverse cysteine-reactive small molecules. Rather than focusing on one particular biological target, we are screening these collections against series of oncologically interesting protein targets using our established method for cysteine-specific proteomic profiling. We term this approach 'chemoproteomics-enabled drug discovery'. Recently, we prepared a collection of chloromethyl triazoles (CMTs) that are accessible from commercially available substrates in just two chemical steps.[17] Using a clickable chloromethyl triazole probe, we performed proteomics analysis to define biomedically interesting targets specific to this chemical scaffold. These targets were then overexpressed and screened against our collection of CMTs *via* gel-based proteomic profiling. For example, we identified compound AA-CW236 as the first non-pseudosubstrate inhibitor of O^6-alkylguanine DNA

Fig. 4. Left: chemical structures of deoxyelephantopin and DEP analog 19a. Right: Predicted model of 19a bound to human PPARγ. This model was calculated using the protein-ligand docking program GOLD.

alkyltransferase (MGMT), a DNA repair protein that renders several devastating cancer forms resistant to chemotherapy (Fig. 5).[18] We confirmed that AA-CW236 potently inhibits MGMT *in vitro* with a K_i value of 24 nM. For comparison, our compound is ~10 times more potent than O[6]-benzylguanine, a known substrate-derived inhibitor of MGMT.[19] Using immunofluorescence microscopy to visualize guanine alkylation, we affirmed that AA-CW236 is also active in cells. We also evaluated the proteome-wide selectivity of AA-CW236 in seven different cancer cell lines and these experiments confirmed that AA-CW236 is not only very potent but also highly selective for MGMT. Moreover, we performed a protein–ligand docking simulation to understand the origin of potency of this inhibitor. The predicted binding mode indicates a key interaction between the trifluoromethoxy group of AA-CW236 and Tyr114 of MGMT (Fig. 5). Indeed, derivatives of AA-CW236 that lack the -CF$_3$ group show significantly lower inhibitory potency.

Conclusion

In summary, we have developed an efficient cysteine-specific proteomic profiling method and have applied this technique to identify biological targets of cysteine-reactive natural products as well as discover a potent and selective chloromethyl triazole inhibitor of MGMT. We firmly believe that we have only just begun to realize the potential of chemical proteomics as means to discover bioactive cysteine-reactive inhibitors and hope to report many more small
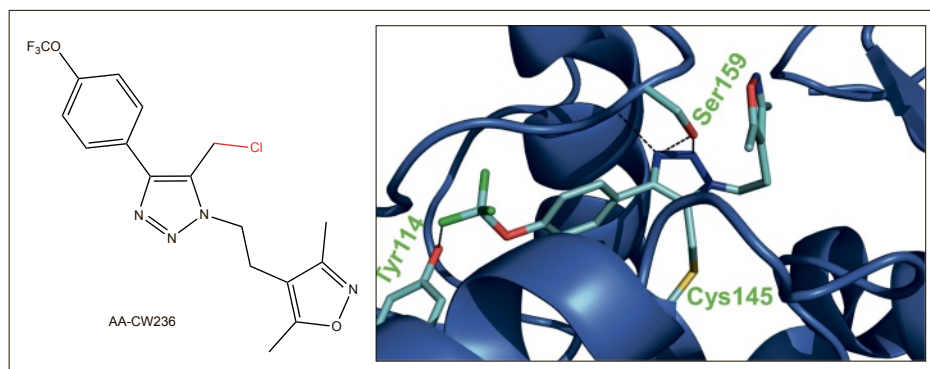

Fig. 5. Left: chemical structure of the chloromethyl triazole inhibitor AA-CW236. Right: predicted model of AA-CW236 bound to human MGMT. This model was calculated using the protein-ligand docking program GOLD.

molecules with exciting biological properties in the near future.

Received: October 3, 2016

[1] E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. D. Dillon, D. A. Bachovchin, K. Mowen, D. Baker, B. F. Cravatt, *Nature* **2010**, *468*, 790.

[2] a) J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discov*. **2011**, *10*, 307; b) R. A. Bauer, *Drug Discov. Today* **2015**, *20*, 1061.

[3] a) M. Schenone, V. Dancik, B. K. Wagner, P. A. Clemons, *Nat. Chem. Biol.* **2013**, *9*, 232; b) S. Ziegler, V. Pries, C. Hedberg, H. Waldmann, *Angew. Chem. Int. Ed.* **2013**, *52*, 2744.

[4] N. Jessani, B. F. Cravatt, *Curr. Opin. Chem. Biol.* **2004**, *8*, 54.

[5] D. A. Shannon, E. Weerapana, *Curr. Opin. Chem. Biol.* **2015**, *24*, 18.

[6] R. Frei, M. D. Wodrich, D. P. Hari, P. A. Bonin, C. Chauvier, J. Waser, *J. Am. Chem. Soc.* **2014**, *136*, 16563.

[7] D. Abegg, R. Frei, L. Cerato, D. P. Hari, C. Wang, J. Waser, A. Adibekian, *Angew. Chem. Int. Ed.* **2015**, *54*, 10852.

[8] B. B. Aggarwal, A. Kumar, A. C. Bharti, *Anticancer Res.* **2003**, *23*, 363.

[9] S. Morgan-Lappe, K. W. Woods, Q. Li, M. G. Anderson, M. E. Schurdak, Y. Luo, V. L. Giranda, S. W. Fesik, J. D. Leverson, *Oncogene* **2006**, *25*, 1340.

[10] G. Lyss, A. Knorre, T. J. Schmidt, H. L. Pahl, I. Merfort, *J. Biol. Chem*. **1998**, *273*, 33508.

[11] J. Lytton, M. Westlin, M. R. Hanley, *J. Biol. Chem*. **1991**, *266*, 17067.

[12] M. X. Su, X. Wu, H. Y. Chung, Y. L. Li, W. C. Ye, *Nat. Prod. Commun.* **2009**, *4*, 1025.

[13] C.-C. Huang, C.-P. Lo, C.-Y. Chiu, L.-F. Shyur, *British J. Pharmacol.* **2010**, *159*, 856.

[14] G. Zou, Z. Gao, J. Wang, Y. Zhang, H. Ding, J. Huang, L. Chen, Y. Guo, H. Jiang, X. Shen, *Biochem. Pharmacol.* **2008**, *75*, 1381.

[15] R. Lagoutte, C. Serba, D. Abegg, D. G. Hoch, A. Adibekian, N. Winssinger, *Nat. Commun.* **2016**, *7*.

[16] Q. Yang, T. E. Graham, N. Mody, F. Preitner, O. D. Peroni, J. M. Zabolotny, K. Kotani, L. Quadro, B. B. Kahn, *Nature* **2005**, *436*, 356.

[17] C. Wang, D. Abegg, D. G. Hoch, A. Adibekian, *Angew. Chem. Int. Ed.* **2016**, *55*, 2911.

[18] M. Esteller, J. G. Herman, *Oncogene* **2004**, *23*, 1.

[19] M. R. Middleton, G. P. Margison, *Lancet Oncol.* **2003**, *4*, 37.