ChemComm

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

ROYAL SOCIETY OF CHEMISTRY View Article Online
DOI: 10.1039/C6CC09555K

Journal Name

COMMUNICATION

Chemoselective and site-selective peptide and native protein modification enabled by aldehyde auto-oxidation

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Landa Purushottam, Srinivasa Rao Adusumalli, Maheshwerreddy Chilamari and Vishal Rai*

We report a chemoselective and site-selective formylation of ε-amine in native proteins. The aldehyde auto-oxidation re-routing, regulated generation of formate, and reversible N-terminus protection drive the transformation. It labels a single ε-amine in a pool of its copies, other nucleophilic residues, and α-amine. The extension of methodology delivers site-selective acylation.

The methods for attachment of synthetic fragments to a protein with precise control over selectivity facilitate myriad applications such as probing post-translational modification, protein design, biophysical investigations and directed therapeutics.[1] A classic route to approach this problem is through pre-engineered proteins[2] with unnatural amino acids amenable to bio-orthogonal chemistry[3] or enzyme-tag pairs.[4] Apparently, these methodologies are not applicable to native protein modification. Chemical methods can potentially meet such technological demands in the years to come.[5] This would require methodologies that can solve two prime challenges, viz. chemoselectivity and site-selectivity (Fig. 1). The prior relates to imparting unique reactivity to a functional group in a protein composed of several nucleophilic amino acid residues. This selected functional group would have multiple copies in a protein where unique reactivity of a single residue would translate the site-selectivity. The differences in the microenvironment of these residues induce a subtle variation in their reactivity. However, its translation into single-site protein labeling for practical applications is rare.[6] Due to these challenges, the prior application of chemical methods is predominantly restricted to proteins that are amenable to tagging via exploitation of single proteinogenic nucleophilic amino acid residue. Besides, the N-terminus α-amine (N^{α}-NH_2) also provides a reactivity hotspot[7] and its modification supersedes that of ε-amine (N^{ε}-NH_2) due to its lower pKa and higher solvent accessibility, in general. The preferential reaction of N^{α}-NH_2 with aldehydes has established the later as a privileged electrophile. The examples include reductive amination,[8] Wittig olefination,[9] Pictet-Spengler ligation,[10] aldehyde capture ligation,[11] imidazolidinone formation,[5a] and transamination.[12] On the other hand, it also highlights the extreme challenge involved in the site-selective N^{ε}-NH_2 modification in preference over N^{α}-NH_2. Even for the proteins devoid of free N-terminus, the high occurrence of Lys residues with similar pKa and solvent accessibility has restrained the identification of N^{ε}-NH_2 with unique reactivity.

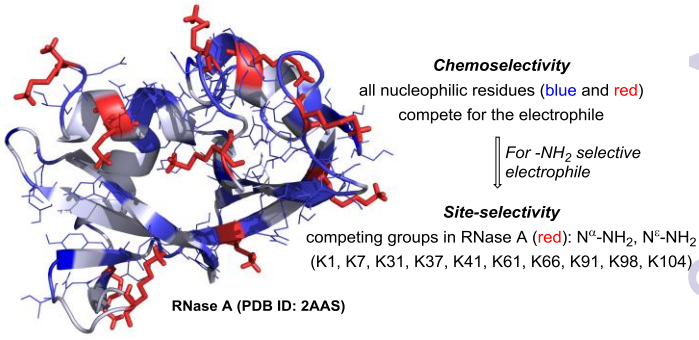


Fig. 1 Challenges in the chemical modification of a native protein.

In a control experiment aimed to understand the reactivity of amine in proteins, we investigated imine formation in physiological conditions (Fig. 2a). The amine in Lys residue and aldehyde react to form electrophilic imine (Fig. 2a, path I) in a reversible reaction. The reaction between a peptide Fmoc-RKH-NH_2 **6a** and benzaldehyde or aromatic aldehydes with 4-Cl, 4-NO_2 and 4-OH substituents (ESI Table S1†) resulted in the formation of Schiff base. However, in one of our experiments, the reaction mixture containing 4-methyl benzaldehyde and **6a** resulted in an adduct that exhibits stability over different modes of chromatography. This compound was isolated (14% yield) and identified as the peptide **7a** with formylated Lys residue (Fig. 2a, path II) after detailed characterization. Here,

Organic and Bioconjugate Chemistry Laboratory, Department of Chemistry Indian Institute of Science Education and Research Bhopal, Bhauri, Bhopal Bypass Road, Bhopal 462066 (India)
E-mail: vrai@iiserb.ac.in

† Electronic Supplementary Information (ESI) available: Additional results and discussion, experimental procedures, characterization data and spectra. See DOI: 10.1039/x0xx00000x

we report this serendipitous finding led chemo- and site-selective N^{ε}-formylation of Lysine. The transformations deliver high levels of efficiency with both peptides and native proteins in physiological conditions with the aldehyde as a formylating pre-reagent. The methodology offers a simple, one-step protocol suitable for modifying single-site of structurally and chemically diverse set of proteins. The additional role of aldehyde in blocking the N-terminus is one of the essential elements for the success of site-selective N^{ε}-formylation. The generality of this technique is demonstrated by the interception of formylation to render N^{ε}-acylation in a simple one-pot process.

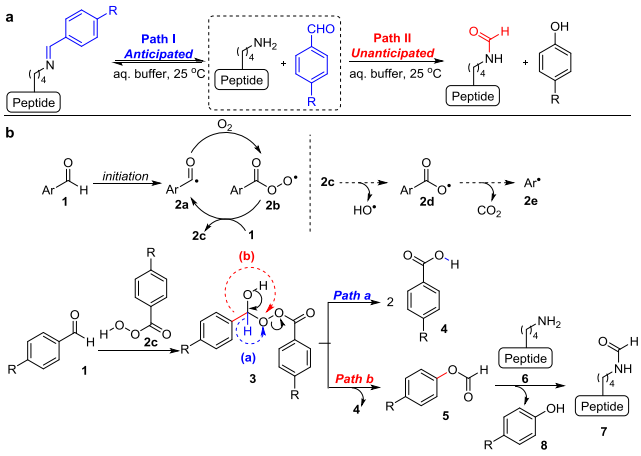


Fig. 2 (a) Imine formation, versus (b) formylation of peptides enabled by aldehyde auto-oxidation.

At first, we evaluated the possibility of formation of product **7a** directly from the aldehyde. However, it required an energetically prohibited C-C bond cleavage in preference to C-N bond (ESI Fig. S1†). In aerobic conditions, the liquid phase oxidation of the aldehyde to carboxylic acids by molecular oxygen is anticipated.[13] However, the control experiments with the carboxylic acid as additive suggested that it does not catalyze the formylation (ESI Fig. S2†). On a closer look, the aerobic oxidation of aldehyde **1** can be initiated by a free radical chain reaction[14] to form the corresponding peracid **2c** (Fig. 2b) identified by ^{1}H NMR spectroscopy (ESI Fig. S3†). Besides, the reaction progress is completely inhibited in the presence of a radical quencher (TEMPO, ESI Fig. S4†). The nucleophilic addition of the peracid **2c** to another molecule of aldehyde **1** generates a tetrahedral adduct **3**. In general, the rearrangement of the tetrahedral adduct **3** proceeds predominantly via migration of the hydrogen to form two molecules of carboxylic acid **4** (Fig. 2b, path a). The formylated peptide **7** indicates that the intermediate **3** from aldehyde **1** has to go through path b to result a molecule each of carboxylic acid **4** and formate **5**. This pathway was predominant for aryl rings with electron donating substituent (ESI Table S1†). Subsequently, the formate ester **5** is trapped by primary amine chemoselectively to result in the formylated peptide **7** (Fig. 2b).

With this information in hand, we altered the substituent from 4-Me to a more electron donating group (4-OMe, ESI Table S1†). It resulted in two folds improvement in conversion (**7a**) stressing on its role in regulating the selectivity between path a and b. With benzaldehyde derivative **1a** bearing 4-$O(CH_2)_3CO_2Et$ substituent, we observed 50% conversion to the formylated product. Next, we hypothesized that hydrophobic groups might regulate the rate of aldehyde oxidation. Such residues can allow access to high effective oxygen concentration at air-water interface through the on-water process.[15] The five folds faster generation of formate **5** in the air in comparison to buffered solution supports the hypothesis (ESI Table S2†). In this purview, the length of alkyl chain was increased from 4-$O(CH_2)_3CH_3$ to 4-$O(CH_2)_9CH_3$ (**1b**). Interestingly, the later aldehyde **1b** resulted in a remarkable increase in conversion to formylated peptide **7a** (45% to 89% conversion). The aldehyde **1b** results in 20% more formylating reagent in comparison to **1a** in physiological conditions. Additionally, the ratio of **5b**:**4b** suggested that path b is followed exclusively with **1b**. For **1a**, the path a competes and dominates over path b (ESI Table S2†). Oxygen is essential for the generation of formate **5** from aldehyde **1b** in absence or presence of phosphate buffer (ESI Table S2†). These investigations allowed understanding of parameters that would regulate the extent and rate of formate (**5**) generation.

The reactive intermediates such as radicals and hydroperoxides[14] (Fig. 2b) have been found to be responsible for the non-selective oxidation of fourteen amino acids in the proteins.[16] In association with epimerization and backbone cleavage, it results in the functional inactivity of proteins. It was imperative to investigate the formation of possible side products through oxidative routes. In this view, we screened the optimized conditions for its efficiency in formylation of N^{ε}-NH_2 in peptides **6**.

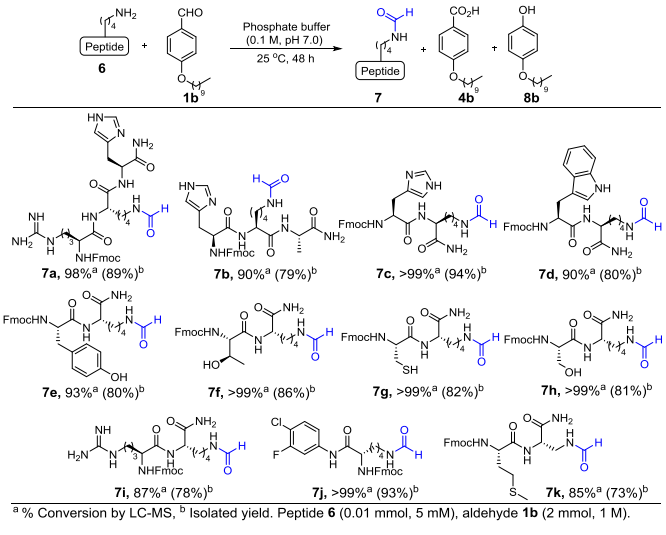


Fig. 3 Chemoselective formylation of peptides **6**.

Under the optimized reaction conditions, no oxidative side products were observed for Ala (**7b**), Arg (**7a**, **7i**), His (**7a**, **7b**,

7c), Trp (**7d**), Tyr (**7e**), Thr (**7f**), Cys (**7g**) and Ser (**7h**) (Fig. 3). However, Met in **6k** undergoes a slow oxidation (~15%) over a period of 48 h. Ambient conditions were found to be ideal for the transformation and necessary for the generation of formate **5**. The backbone of formylated peptides (**7a-k**) were found to be stable and exhibited no epimerization or fragmentation (Fig. 3). Moreover, it was exciting to note that the formylated peptide was isolated in excellent yields in all the cases (**7a-k**). The method can serve as an efficient chemoselective route to access formylated primary amines in peptides. It can also offer activity based probes for readers of formylation, a post-translational modification.[17] The regulation of formate generation was critical for avoiding the interference with competitive pathways and success of methodology. Two aldehydes (**1a** and **1b**) offered desired rate of reaction retaining the selectivity and were screened for their application in formylation.

In the next step, we examined the efficiency of this protocol with proteins. RNase A **9a** served as a model protein. It contains ten lysine residues, where nine of them are solvent exposed. We were excited to note the formation of monolabeled RNase A **10f** (82% conversion, Fig. 5) upon vortexing **9a** with aldehyde **1b**. The N^{ε}-K37 is labeled site-selectively and homogeneously with no traces of N^{α}-NH_2 modification (MS and MS^2). The methodology distinguishes a single ε-amine (K37) from nine other Lys residues.

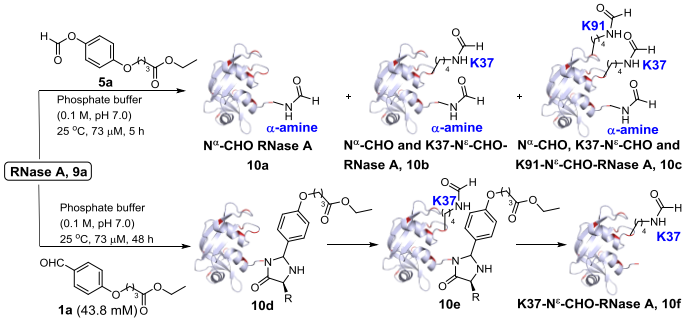


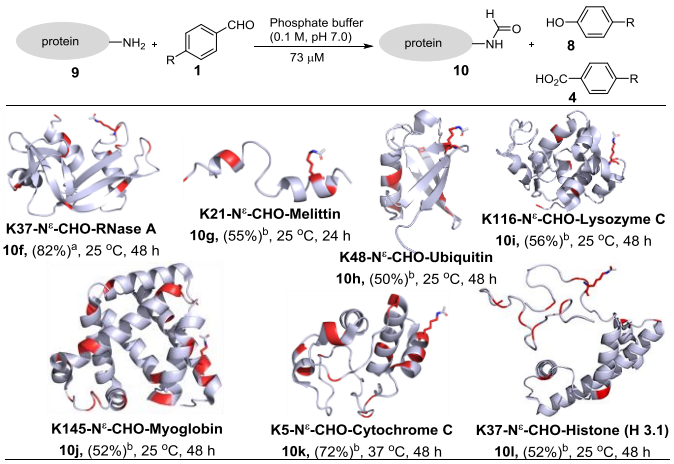
Fig. 4 Insight to the site-selectivity of formylation.

The two roles played by aldehyde are important to achieve the observed dual selectivity i.e. chemo- and site-selectivity. In the first step, α-amine reacts rapidly with an aldehyde that serves as a transient protecting group. Subsequently, formate generated from aldehyde results in the formylation of N^{ε}-K37. At this point, we argued that the high reactivity of a particular N^{ε}-NH_2 would mean drift of its pKa towards that of N^{α}-NH_2. In such a case, both will render a similar reactivity profile. The selective protection of N^{α}-NH_2 (**10d**, Fig. 4) suggests that the reversible formation of imidazolidinone (M+218, ESI Fig. S6†), and not imine, is the key contributor for rendering it unreactive. On the contrary, imine from N^{ε}-NH_2 lacks suitably positioned amide for an intramolecular cyclization.

To further substantiate the origin of selectivity, we carried out a list of control experiments. The overall concentration of protein in the reaction milieu was found to be important to avoid protein self-assembly associated irreproducibility.[18] The pre-synthesized formate ester **5a** (5 equiv.) was vortexed with RNase A at 25 °C. The stoichiometry and rate of addition of **5a** were acquired from its rate of formation from aldehyde **1a** (ESI Fig. S7†). Under the given conditions, a mixture of mono-, bis- and tris- formylated RNase A **10a-10c** was observed in 5 h (Fig. 4). The N^{α}-NH_2 modification is prevalent reasserting its preferential reactivity followed by the labeling of K37-N^{ε}-NH_2 and then K91-N^{ε}-NH_2. This substantiates that K37-N^{ε}-NH_2 is the most reactive Lys residue in RNase A **9a**. In parallel, RNase A **9a** and formate **5a** were incubated in the presence of pyridine 2-carboxaldehyde (ESI Fig. S8†). In this case, the imidazolidinone[5a] forms at N^{α}-NH_2, but the formyl transfer results in a mixture of mono-, bis-, and tris-labeled RNase A.

After establishing the methodology with RNase A, we extended the protocol to other proteins. Melittin, a 27 amino acid long protein, bears three solvent exposed Lys residues. The formylating pre-reagent **1a** displayed a good control over selectivity in this case. It labels the K21-N^{ε}-NH_2 with absolute chemo- and site-selectivity (**10g**, Fig. 5). Direct formylation of Melittin with pre-synthesized formate ester **5a** resulted in a heterogeneous mixture of mono-, bis- and tris-labeled Melittin involving N^{α}-NH_2, K21-N^{ε}-NH_2, and K23-N^{ε}-NH_2. The most reactive ε-amine is K21 residue in this case. At this point, we decided to examine the efficiency of protocol with Ubiquitin that bears seven Lys residues. The control experiment confirmed N^{α}-NH_2 to be the most reactive amine followed by K48-N^{ε}-NH_2 and K63-N^{ε}-NH_2. This reaction is unable to distinguish the reactivity order of the latter two amines. However, K48-N^{ε}-NH_2 is labeled with exclusive selectivity (**10h**) when Ubiquitin is vortexed with **1a**. This example showcases the strength of the protocol in distinguishing subtle differences in the reactivity of Lys residues. We further extended the methodology for Lysozyme C (six Lys) modification. The formylation of K116-N^{ε}-NH_2 progressed with absolute site-selectivity and resulted in homogeneously labeled protein (**10i**). It is interesting to note that the kinetic reactivity order of amines suggested by calculated solvent accessibility ranks K116-N^{ε}-NH_2 at the third position.[7b] However, our control experiments through pre-generated formate ester confirm that K116 is the most reactive primary amine. In Myoglobin, K145-N^{ε}-NH_2 is more reactive than N^{α}-NH_2 and eighteen other Lys residues. It is also the exclusive site of modification (**10j**) upon reaction with **1b** as formylating pre-reagent. Subsequently, a Lys-rich (19 copies) protein Cytochrome C with blocked N-terminus (N^{α}-$NHCOCH_3$) was subjected to formylation. In this case, K5-N^{ε}-NH_2 underwent formylation and resulted in homogenously monolabeled protein **10k**. Finally, we selected histone (H 3.1) that has been one of the prime targets for adventitious post-translational formylation in human nuclear proteins.[19] Here, site-selective labeling of K37-N^{ε}-NH_2 resulted in homogeneously labeled **10l**. It was interesting to note that the formylated Lys residues are part of α-helix (**10g**, **10j**, **10k**), loop (**10f**, **10i, 10l**) as well as β-sheet (**10h**). The structure of proteins **10f**-**10k** remains unaltered post-formylation as suggested by the circular dichroism spectroscopy (ESI Fig. S5†). The enzymatic activity of

formylated enzymes **10f** and **10i** is conserved and highlights the mild nature of reaction conditions (ESI Fig. S37 and S38†).



[a] % Conversion by LC-MS. **1b** (43. 8 mM) used as formylating pre-reagent. Site of formylation remains unaltered with **1a.** [b] % Conversion by MALDI. **1a** (43.8 mM) is formylating pre-reagent. For Melittin, **1a** (21.9 mM) is formylating pre-reagent. The reactions were performed in triplicate.

Fig. 5 Site-selective native protein modification enabled by auto-oxidation of aldehydes.

In the pursuit to examine the generality of N^{ε}-NH_2 labeling protocol, we designed an experiment for interception of formylation by an acylating reagent (Fig. 6). It was exciting to note that appropriate selection of electrophile **11** leads to site-selective and homogeneous acylation of RNase A. There is no indication of formylation or oxidative pathway under the optimized conditions. The labeled protein K37-N^{ε}-COR_1-RNase A (**10m**) is primed with maleimide and offers a bio-orthogonal site for late-stage modification with the desired tag of interest. For example, the reaction of **10m** with 7-mercapto-4-methylcoumarin **12** delivers **13** efficiently. To gain insight to acylation, we performed the reaction in the absence of aldehyde **1a**, keeping the other parameters constant (ESI Fig. S9†). It results in a mixture of mono-, bis- and tris-labeled products. We expect the labeling of N^{α}-NH_2 and K37-N^{ε}-NH_2 in the acylation. However, the labeling of an additional N^{ε}-NH_2 suggests that the reduced effective concentration of free Lys residues, ascribed to imine formation in the presence of aldehyde **1a**, has a role in regulating the site-selectivity.

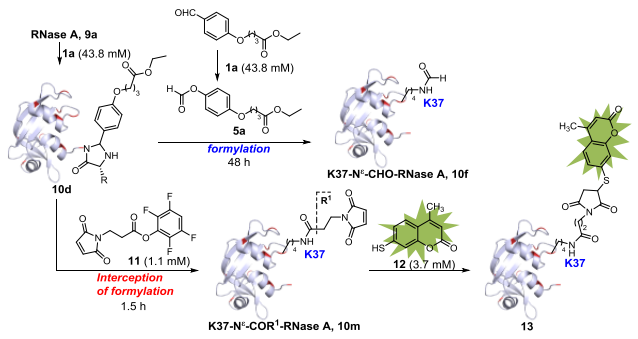


Fig. 6 Interception of formylation renders site-selective acylation (ESI Fig. S10†) and offers late-stage modification.

In summary, we have successfully demonstrated that a subtle difference in reactivity of native protein backbone residues can be distinguished through a chemical transformation to provide a platform for site-selective labeling. The protocol renders access to single site formylation of ε-amine in native proteins, a recently identified post-translational modification.[17] It is noteworthy that the aldehyde auto-oxidation generates formylating pre-reagent in physiological conditions. Contrary to the general belief, the free radicals involved in the process demonstrate mild nature and do not promote alternative reaction pathways. The site-selective reversible protection of N^{α}-NH_2, regulation of formate generation and effective N^{ε}-NH_2 concentration serves as the key to one-pot site-selective N^{ε}-NH_2 labeling. Even though a complex combination of multiple parameters is involved, operationally simple front-end protocol would also excite the non-experts. The transformation is efficient in the modification of peptides and structurally diverse proteins with a wide range of Lys residues (3-19). Interception of formylation of protein by a distinct electrophile confirms extended application of this chemical platform. We believe that the demonstration of unique reactivity and kinetic selectivity of protein side chain residue delivers a breakthrough that will fuel the ongoing efforts for site-selective modification of native proteins.

View Article Online DOI: 10.1039/C6CC09555K

Notes and references

1 N. Krall, F. P. da Cruz, O. Boutureira and G. J. L. Bernardes, *Nat. Chem*., 2016, **8**, 103.
2 C. H. Kim, J. Y. Axup and P. G. Schultz, *Curr. Opin. Chem. Biol*., 2013, **17**, 412.
3 C. D. Spicer and B. G. Davis, *Nat. Commun*., 2014, **5**, 4740.
4 D. Rabuka, *Curr. Opin. Chem. Biol*., 2010, **14**, 790.
5 (a) J. I. MacDonald, H. K. Munch, T. Moore and M. B. Francis, *Nat. Chem. Biol*., 2015, **11**, 326; (b) Y. Takaoka, A. Ojida and I. Hamachi, *Angew. Chem., Int. Ed*., 2013, **52**, 4088.
6 J. M. McFarland and M. B. Francis, *J. Am. Chem. Soc.*, 2005, **127**, 13490.
7 (a) R. Singudas, S. R. Adusumalli, P. N. Joshi and V. Rai, *Chem. Commun*., 2015, **51**, 473; (b) X. Chen, K. Muthoosamy, A. Pfisterer, B. Neumann and T. Weil, *Bioconjugate Chem*., 2012, **23**, 500; (c) A. O.-Y. Chan, C.-M. Ho, H.-M. Chong, Y.-C. Leung, J.-S. Huang, M.-K. Wong and C.-M. Che, *J. Am. Chem. Soc.,* 2012, **134**, 2589.
8 J. C. Gildersleeve, O. Oyelaran, J. T. Simpson and B. Allred, *Bioconjugate Chem*., 2008, **19**, 1485.
9 K. M. Lum, V. J. Xavier,M. J.-H. Ong, C. W. Johannes and K.-P. Chan, *Chem. Commun*., 2013, **49**, 11188.
10 P. Agarwal, J. van der Weijden, E. M. Sletten, D. Rabuka and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A*., 2013, **110**, 46.
11 M. Raj, H. Wu, S. L. Blosser, M. A. Vittoria and P. S. Arora, *J. Am. Chem. Soc*., 2015, **137**, 6932.
12 L. S. Witus, C. Netirojjanakul, K. S. Palla, E. M. Muehl, C. Weng, A. T. Iavarone and M. B. Francis, *J. Am. Chem. Soc*., 2013, **135**, 17223.
13 J. R. McNesby and Jr. C. A. Heller, *Chem. Rev*., 1954, **54**, 325.
14 V. Chudasama, R. J. Fitzmaurice and S. Caddick, *Nat. Chem.,* 2010, **2**, 592.
15 R. N. Butler and A. G. Coyne, *Chem. Rev*., 2010, **110**, 6302.
16 G. Xu and M. R. Chance, *Chem. Rev*., 2007, **107**, 3514.
17 T. Wang, Q. Zhou, F. Li, Y. Yu, X. Yin and J. Wang, *ChemBioChem.,* 2015, **16**, 1440.
18 P. N. Joshi, P. Landa, N. K. Das, S. Mukherjee and V. Rai, *RSC Adv*., 2016, **6**, 208.
19 T. Jiang, X. Zhou, K. Taghizadeh, M. Dong and P. C. Dedon, *Proc. Natl. Acad. Sci. U.S.A.,* 2007, **104**, 60.