Protein Labeling

International Edition: DOI: 10.1002/anie.201502618
German Edition: DOI: 10.1002/ange.201502618

Covalent Protein Labeling by Enzymatic Phosphocholination**

Katharina Heller, Philipp Ochtrop, Michael F. Albers, Florian B. Zauner, Aymelt Itzen, and Christian Hedberg**

Dedicated to Professor Roger S. Goody on the occasion of his 71th birthday

Abstract: *We present a new protein labeling method based on the covalent enzymatic phosphocholination of a specific octapeptide amino acid sequence in intact proteins. The bacterial enzyme AnkX from Legionella pneumophila has been established to transfer functional phosphocholine moieties from synthetically produced CDP-choline derivatives to N-termini, C-termini, and internal loop regions in proteins of interest. Furthermore, the covalent modification can be hydrolytically removed by the action of the Legionella enzyme Lem3. Only a short peptide sequence (eight amino acids) is required for efficient protein labeling and a small linker group (PEG-phosphocholine) is introduced to attach the conjugated cargo.*

Site-directed labeling strategies are essential for modifying proteins with functionalities not defined by the genetic code. It is generally desirable to introduce modifications by regioselective chemical reactions to maintain protein activity and integrity with minimal cross-reactivity.[1] An alternative approach is the utilization of enzymes that specifically recognize small amino acid sequences in a protein of interest.[2] For instance, phosphopantetheinyl (Ppant) transferase (PPTase) physiologically transfers a Ppant moiety from an activated precursor nucleotide (coenzyme A) to a serine of the undecapeptide sequence DSLEFIASKLA.[3] The Ppant is chemically modified and combined with the PPTase-catalyzed reaction to transfer functional Ppant to recombinantly produced proteins containing the PPTase recognition sequence.[3b,c,4] If, however, the recognition sequence is replaced by the 8 kDa acyl carrier protein (ACP), modification and subsequent cleavage with PPTase and ACP hydrolase (AcpH), respectively, is achievable.[4b] However, recently the demodification of the 11 amino acid recognition sequence by different AcpH has been reported.[5] We sought to develop an alternative enzymatic protein labeling strategy utilizing a defined nucleotide that would make it possible to conveniently attach and detach a label of choice at a short peptide sequence. Here, we present data on the general applicability of covalent enzymatic phosphocholination of proteins for regioselective labeling and delabeling using synthetic CDP-choline derivatives.

The recently discovered host cell phosphocholination by *Legionella pneumophila* is a posttranslational modification in which a phosphocholine moiety is enzymatically transferred via the Legionella effector protein AnkX from a cytidine diphosphate choline (CDP-choline) to a serine residue in the switch II loop of the small GTPase Rab1 in the host cell (Scheme 1 A).[6] Interestingly, the Rab1 phosphocholine modification is hydrolytically cleaved by the Legionella phosphodiesterase Lem3 at a later stage of infection (Scheme 1 A).[7] We have recently observed that AnkX recognizes only the TITSSYYR peptide sequence of the switch II loop of Rab1b, but does not necessarily require the discrete GTPase structure.[7a]

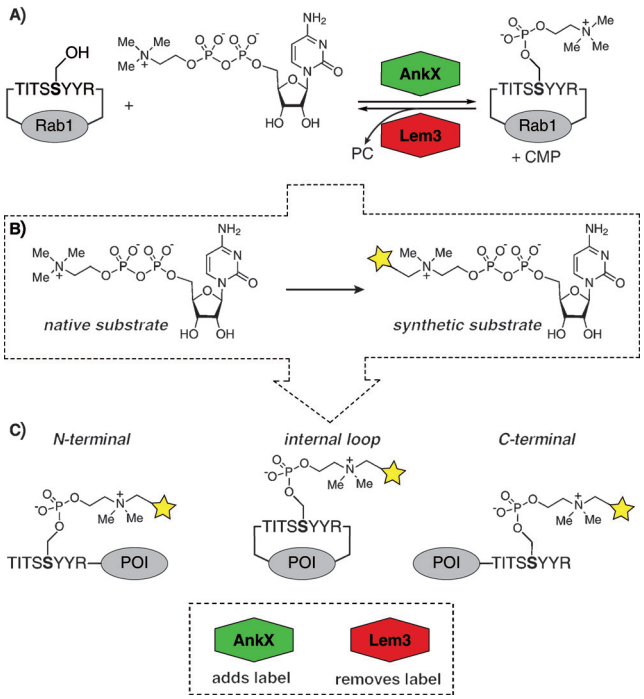
We envisioned modification/synthesis of functionalized CDP-choline derivatives carrying any label of choice (Scheme 1 B), which could be transferred to fusion proteins of interest, where the recognition sequence could be located either at N-terminal, C-terminal, or in internal loop regions (Scheme 1 C).

First, we investigated several AnkX (949 amino acids)[6a,8] and Lem3 (570 amino acids)[7] constructs for their performance in protein expression and enzymatic activity (Figure S1). Constructs $AnkX_{1–800}$ and full-length Lem3 were found to be superior with respect to expression, purification, and desired activity, and were therefore retained for establishing our protein labeling strategy. In an alanine scan of the AnkX recognition sequence, we established that modifications up- and downstream of octapeptide TITS*S*YYR were fully accepted, and modifications (marked in bold) within **TI**TS-*S*YY**R** attenuated acceptor capability. Changes within TS*S*YY

[*] M.Sc. K. Heller,[+] F. B. Zauner, Prof. Dr. A. Itzen
Center for Integrated Protein Science Munich
Technische Universität München, Department Chemistry
Lichtenbergstrasse 4, 85748 Garching (Germany)
E-mail: aymelt.itzen@tum.de
M.Sc. P. Ochtrop,[+] M.Sc. M. F. Albers, Prof. Dr. C. Hedberg
Chemical Biology Center (KBC), Institute of Chemistry
Umeå University, 90187 Umeå (Sweden)
E-mail: christian.hedberg@umu.se
Prof. Dr. C. Hedberg
Max Planck Institute of Molecular Physiology
Department of Chemical Biology
Otto-Hahn-Strasse 11, 44227 Dortmund (Germany)
[+] These authors contributed equally to this work.

[**] This work was supported by a priority program (SPP 1623) and a collaborative research center (SFB 1035, project B05) of the German Research Foundation (DFG) as well as by the Knut and Alice Wallenberg Foundation (KAW 2013.0187), Sweden. M.F.A. thanks the Max Planck Society for a fellowship from the International Max Planck Research School (IMPRS). We are indebted to Sascha Gentz, Dr. Matthias Müller (MPI-Dortmund), Prof. Stephan Sieber, Prof. Johannes Buchner, Dr. Martin Haslbeck, and Dr. Sabine Schneider (TU Munich) for useful discussions and the provision of materials.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201502618.



Scheme 1. A) Protein phosphocholination and dephosphocholination by the Legionella enzymes AnkX and Lem3. B) Design of a CDP-choline derivative bearing a label at one of the methyl positions on the quaternary amine. C) Design of protein substrates (protein of interest, POI) equipped with the octapeptide recognition sequence at N- and C-termini and at internal loop regions.

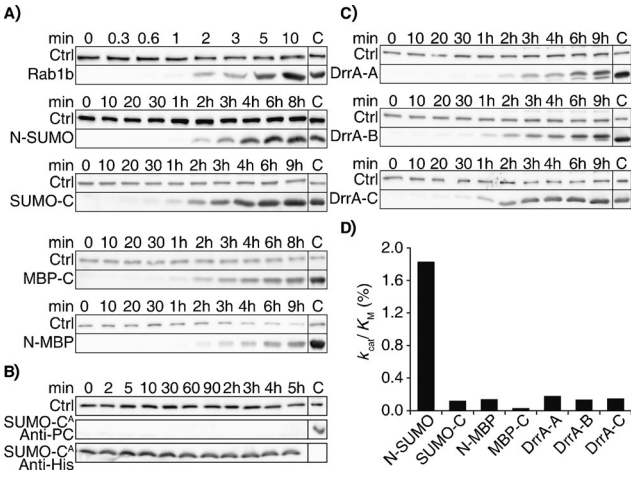


Figure 1. Phosphocholination of AnkX model substrates. A) Rab1b$_{3–174}$ and various protein model substrates (SUMO, MBP) equipped with the octapeptide recognition sequence at N- and C-termini are monitored by western blotting against phosphocholine (for analysis see Figure S6). B) Negative control corresponding to Figure 1 A with SUMO-TITS**A**YYR. Anti-His$_6$ blotting served as a loading control. C) Enzymatic phosphocholination of DrrA$_{340–533}$ at internal loop regions that have been extended with the AnkX octapeptide recognition motif, monitored by western blotting against phosphocholine. D) Catalytic efficiencies (k_{cat}/K_M) of the phosphocholination reaction on the various protein substrates relative to the corresponding reaction on the native substrate Rab1b$_{3–174}$. Ctrl: loading control of individual experiments with His$_6$-MBP-PC-Rab1b$_{3–174}$; C: intensity of quantitatively phosphocholinated protein.

completely abolished phosphocholination, suggesting that the octapeptide sequence TITS*S*YYR is the minimum conserved recognition motif for efficient phosphocholination by AnkX (Figure S2). In order to investigate whether the octapeptide sequence could be phosphocholinated by AnkX$_{1–800}$ as protein fusions, we recombinantly attached the TITSSYYR octapeptide sequence at the N- and C-termini of the maltose binding protein (MBP) and the small ubiquitin-like modifier (SUMO) and carried out heterologous expression in *E. coli* (Figure S3 and Supporting Information). Phosphocholination of the model proteins (50 μM) was investigated with a catalytic amount of AnkX (1 μM) in the presence of 1 mM CDP-choline and analyzed by ESI-MS after 24 h of incubation (Figure S4). SUMO constructs (TITSSYYR-SUMO and SUMO-TITSSYYR) were phosphocholinated at the N- and C-termini (Figure S4) and MBP at its N-terminus (but also minor modification of C-terminally tagged MBP), verifying that phosphocholination can also be introduced into model proteins. The modifications of the AnkX substrates were validated independently by detecting protein phosphocholination with an anti-phosphocholine antibody by western blotting (Figure 1 A–C). In this assay, also phosphocholination of MBP at the C-terminus was detected. By using the identical assay we could show that phosphocholine transfer is dependent on the correct peptide sequence and acceptor amino acid (Figure 1 A), since the SUMO-TITS*A*YYR substrate (lacking the serine that is targeted by AnkX) failed to undergo AnkX-mediated phosphocholination (Figure 1 B). Including AnkX recognition motifs into internal loop regions of target proteins could significantly expand the application of protein labeling by phosphocholination. We therefore introduced the TITSSYYR octapeptide sequence into the GEF domain of the enzyme DrrA (amino acids 340–533, DrrA$_{340–533}$)[9] (Figure S3): 1) E$_{425}$-TITSSYYR-S$_{426}$ (DrrA-A), 2) T$_{455}$-TITSSYYR-P$_{456}$ (DrrA-B), 3) N$_{509}$-TITSSYYR-V$_{510}$ (DrrA-C).

Next, we investigated the phosphocholination by AnkX$_{1–800}$ of these proteins via Western blotting using the anti-phosphocholine antibody (Figure 1 C) and ESI-MS (Figure S4). The phosphocholination of internal loop regions on DrrA was readily detected, thus suggesting that the AnkX recognition motif TITSSYYR can be used for phosphocholination in internal loops. Full enzymatic phosphocholination of peptide-tagged SUMO, MBP, and DrrA proteins could be achieved with extended reaction times, thus demonstrating that the enzymatic reaction does not halt at an intermediary labeling state (Figure S4). Finally, we derived time-course data of the AnkX phosphocholination reaction using quantitative densitometry of the western blots. The apparent catalytic efficiencies (k_{cat}/K_M) of the reaction were calculated to compare the different AnkX substrates (Figure 1 D and Figures S5 and S6). Not surprisingly, the artificial substrates (i.e. peptide-tagged SUMO, MBP, and DrrA) perform less well than the native Rab1b protein substrate, which may be attributed to increased conformational flexibility (e.g. SUMO, MBP) and/or structural restrictions (e.g. internal loops of DrrA).

The development of phosphocholination by AnkX as a protein labeling strategy depends on the synthesis and

utilization of CDP-choline derivatives permitting the transfer of the label along with the derivatized phosphocholine group. The crystal structure of $AnkX_{1–484}$ in complex with CDP-choline shows the quaternary choline ammonium group positioned in a deep, but solvent-accessible pocket,[8] thus suggesting that any label should be connected to the quaternary ammonium function by a flexible hydrophilic linker of sufficient length (Figure S7). We generated synthetic CDP-choline derivatives containing a fluorescein label connected to the amine group via a polyethylene glycol (PEG) linker of varying length (**1**–**5**, Figure 2A) (for synthesis details, see Figure S7).[10]

Next, we investigated qualitatively whether the CDP-choline derivatives **1**–**5** are substrates for AnkX (Figure 2B). Rab1b was used as the acceptor substrate (50 μM) and phosphocholination was monitored after the addition of the CDP-choline derivatives (1 mM) in the presence of catalytic amounts of $AnkX_{1–800}$ (0.5 μM) by ESI-MS (Figure S8) and in-gel fluorescence after SDS-PAGE (Figure 2B). First, we probed the active site of AnkX with fluorescent derivatives **1**–**3**; **1** and **2** were transferred by $AnkX_{1–800}$, but not derivative **3**, where the ammonium/amine function was replaced by oxygen, suggesting the importance of the positive charge for substrate recognition. This is in line with the observation that the choline group forms salt bridges with E_{226} and D_{265} and a cation–π interaction with F_{107} of AnkX.[8] Next, we focused on the PEG spacer using derivatives **1**, **4**, and **5**, decreasing the spacer length stepwise from six PEG units to only two. To our surprise, derivatives **1** (PEG_6) and **4** (PEG_4) were transferred with an apparently similar efficiency. Only derivative **5** (PEG_2), where the spacer clearly is too short, was not transferred. Conveniently, the attachment of PC-PEG_x-fluorescein moieties to the protein substrates resulted in a pronounced shift in electrophoretic mobility and thus permitted quantification of its incorporation using gel densitometry of the Coomassie-stained SDS-PAGE bands. Both **1** and **4** were quantitatively transferred to Rab1b, indicating that full modification of protein substrates is achieved using AnkX-mediated phosphocholine transfer. Nucleotide **2** was a worse substrate and modified only roughly 50% of the protein in 240 min. We also obtained apparent catalytic efficiencies (k_{cat}/K_M) from the time course experiments (Figure 2B,D). Here, substrate **4** (PEG_4) performed significantly better than **1** (PEG_6), indicating that further optimization of the linker length and linker chemistry could lead to additional improvement of the substrate properties of CDP-choline derivatives.

In addition to Rab1b, we investigated the transfer of CDP-choline derivatives **1**, **2**, and **4** to the TITSSYYR-SUMO construct by detecting modified proteins with ESI-MS and in-gel fluorescence (SDS-PAGE) (Figure 2C). The substrate profile was identical to that of Rab1b, thus confirming that the protein modification is dependent only on the presence of the AnkX-recognition sequence TITSSYYR. However, the extent of PC-PEG_x-fluorescein attachment to TITSSYYR-SUMO differed among the nucleotide substrates. The band shift corresponding to the attachment of **1** indicated only a slight incorporation into the proteins, whereas **4** modified approximately 70% of TITSSYYR-SUMO in 180 min (Figure 2C). This is not surprising since the native AnkX substrate

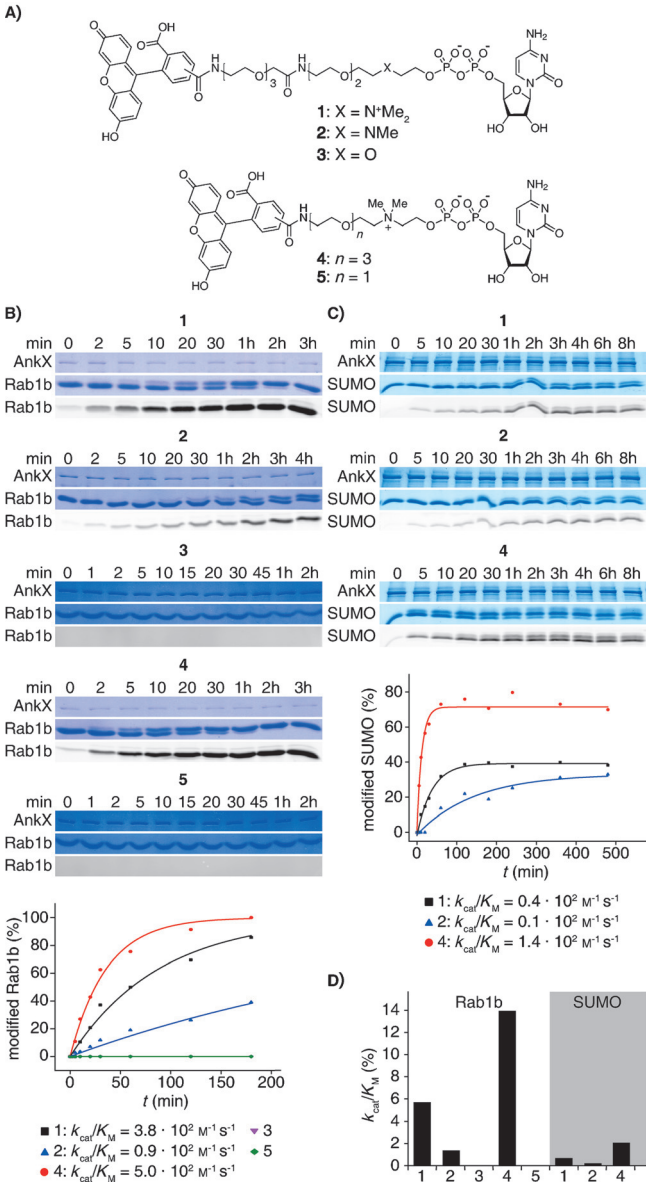


Figure 2. AnkX-catalyzed modification using CDP-choline derivatives. A) Structures of synthesized CDP-choline derivatives carrying PEG-fluorescein and containing different functional groups (**1**: quaternary amine, **2**: tertiary amine, **3**: oxygen) or having different PEG linker lengths (**1**, **4**, **5**). Fluorescein labeling of $Rab1b_{3–174}$ (B) and TITSSYYR-SUMO (C) with CDP-choline derivatives. The time course of phosphocholination was monitored through in-gel fluorescence of SDS-PAGE gels. Successful phosphocholination was dependent on the linker length (**1**, **4**) and the presence of positive charge at the position of the quaternary amine (**1**, **2**, **4**). D) Catalytic efficiencies for the modification reaction with the various nucleotide derivatives were plotted relative to the native reaction.

Rab1b is a more efficient recipient of the modification than the artificially generated peptide substrates. Still, the combination of TITSSYYR-tagged proteins with CDP-choline derivatives (e.g. **4**) allows for efficient labeling using AnkX enzyme on a reasonable timescale.

Next we asked whether Lem3 could be utilized to detach functional phosphocholine derivatives from modified target proteins and whether the peptide sequence is important for

recognition by Lem3.[7,11] Kinetic investigations revealed that alanine substitutions except for the first threonine in the octapeptide sequence (TITSSYYR) affected the Lem3 activity by 10–50% (Figure S9). We subjected fluorescently phosphocholinated $Rab1b_{3-174}$ and TITSSYYR-SUMO (modified with constructs **1**, **2**, and **4**) to Lem3-mediated enzymatic dephosphocholination. In-gel fluorescence after SDS-PAGE, and differences in the electrophoretic mobility of modified versus unmodified proteins (Figure 3 A,B) revealed that $Rab1b_{3-174}$ and TITSSYYR-SUMO could be effectively dephosphocholinated by Lem3. It appears like the internal loop region of Rab1b is the most efficient substrate for Lem3, although TITSSYYR-SUMO reacts on a comparable timescale. This is reflected in similar catalytic efficiencies of Lem3 for the different PC-PEG$_x$-fluorescein modified protein substrates (Figure 3 A,B).

In summary, we present a covalent enzymatic protein labeling and delabeling method based on a small peptide recognition motif. The labeling method is applicable to N-terminal, C-terminal, and internal loop recognition motifs of only eight amino acids in proteins of interest. The recognition sequence is small compared to those of other labeling techniques that require N- or C-terminal attachment of entire protein domains (e.g. ACP,[4b] fluorescent proteins,[12] SNAP-tag,[13] CLIP-tag,[14] etc.) also permitting internal labeling. The method is fully compatible with current existing chemical labeling methodologies, since the recognition sequence contains no cysteine or other reactive amino acids and therefore can be used easily for double labeling approaches. The possibility of conveniently detaching labels using Lem3-mediated hydrolysis makes it possible to recover protein samples and use them subsequently for another labeling strategy. Given the importance of selective protein labeling methods in biosciences, enzymatic phosphocholination for protein modification holds the potential to increase the scope of possible labeling strategies in complex biological systems.

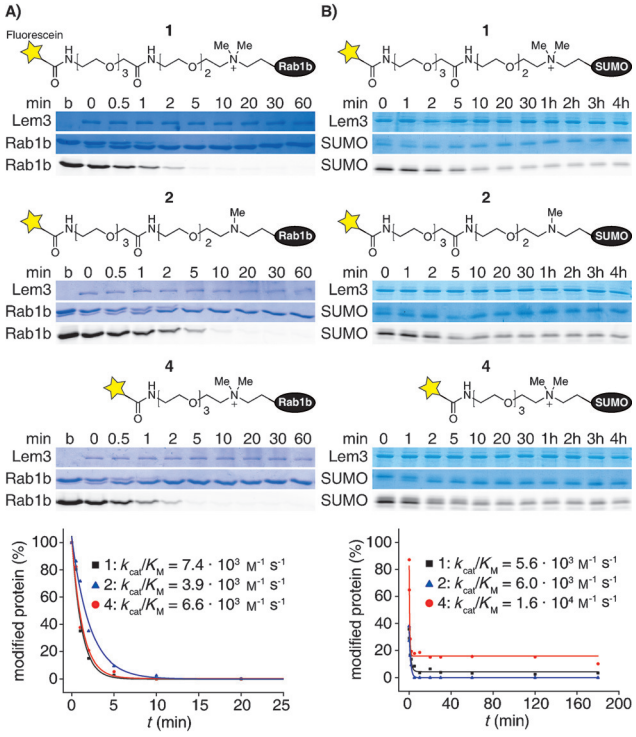


Figure 3. Enzymatic dephosphocholination by Lem3. Lem3 dephosphocholination of protein substrates $Rab1b_{3-174}$ (A) and TITSSYYR-SUMO (B) preparatively modified with fluorescent phosphocholine derivatives, analyzed by in-gel fluorescence of SDS-PAGE (gray scale) and Coomassie staining (blue). Yellow star: fluorescein.

Keywords: enzymes · nucleotides · phosphocholination · protein modifications

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 10327–10330
Angew. Chem. **2015**, *127*, 10467–10471

[1] a) D. M. Patterson, L. A. Nazarova, J. A. Prescher, *ACS Chem. Biol.* **2014**, *9*, 592–605; b) E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998; *Angew. Chem.* **2009**, *121*, 7108–7133.
[2] M. Rashidian, J. K. Dozier, M. D. Distefano, *Bioconjugate Chem.* **2013**, *24*, 1277–1294.
[3] a) R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem. Biol.* **1996**, *3*, 923–936; b) J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15815–15820; c) Z. Zhou, P. Cironi, A. J. Lin, Y. Q. Xu, S. Hrvatin, D. E. Golan, P. A. Silver, C. T. Walsh, J. Yin, *ACS Chem. Biol.* **2007**, *2*, 337–346; d) Z. Zhou, A. Koglin, Y. Wang, A. P. McMahon, C. T. Walsh, *J. Am. Chem. Soc.* **2008**, *130*, 9925–9930.
[4] a) J. Yin, A. J. Lin, D. E. Golan, C. T. Walsh, *Nat. Protoc.* **2006**, *1*, 280–285; b) N. M. Kosa, R. W. Haushalter, A. R. Smith, M. D. Burkart, *Nat. Methods* **2012**, *9*, 981–984.
[5] N. M. Kosa, K. M. Pham, M. D. Burkart, *Chem. Sci.* **2014**, *5*, 1179–1186.
[6] a) S. Mukherjee, X. Y. Liu, K. Arasaki, J. McDonough, J. E. Galan, C. R. Roy, *Nature* **2011**, *477*, 103–106; b) C. A. Hardiman, C. R. Roy, *Mbio* **2014**, *5*, e01035-13; c) X. X. Pan, A. Lührmann, A. Satoh, M. A. Laskowski-Arce, C. R. Roy, *Science* **2008**, *320*, 1651–1654; d) M. P. Müller, H. Peters, J. Blümer, W. Blankenfeldt, R. S. Goody, A. Itzen, *Science* **2010**, *329*, 946–949.
[7] a) P. R. Goody, K. Heller, L. K. Oesterlin, M. P. Müller, A. Itzen, R. S. Goody, *EMBO J.* **2012**, *31*, 1774–1784; b) Y. H. Tan, R. J. Arnold, Z. Q. Luo, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 21212–21217.
[8] V. Campanacci, S. Mukherjee, C. R. Roy, J. Cherfils, *EMBO J.* **2013**, *32*, 1469–1477.
[9] S. Schoebel, L. K. Oesterlin, W. Blankenfeldt, R. S. Goody, A. Itzen, *Mol. Cell* **2009**, *36*, 1060–1072.
[10] a) C. M. Crane, J. Kaiser, N. L. Ramsden, S. Lauw, F. Rohdich, W. Eisenreich, W. N. Hunter, A. Bacher, F. Diederich, *Angew. Chem. Int. Ed.* **2006**, *45*, 1069–1074; *Angew. Chem.* **2006**, *118*, 1082–1087; b) J. G. Moffatt, H. G. Khorana, *J. Am. Chem. Soc.* **1961**, *83*, 649–658.
[11] M. F. Albers, C. Hedberg, *J. Org. Chem.* **2013**, *78*, 2715–2719.
[12] N. C. Shaner, P. A. Steinbach, R. Y. Tsien, *Nat. Methods* **2005**, *2*, 905–909.
[13] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2002**, *21*, 86–89.
[14] A. Gautier, A. Juillerat, C. Heinis, I. R. Correa, M. Kindermann, F. Beaufils, K. Johnsson, *Chem. Biol.* **2008**, *15*, 128–136.

Received: March 21, 2015
Published online: July 3, 2015