A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition Chemie

GDCh

www.angewandte.org

Accepted Article

Title: The First "Zero-Length" Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis

Authors: Christoph Hage, Claudio Iacobucci, Anne Rehkamp, Christian Arlt, and Andrea Sinz

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.201708273
Angew. Chem. 10.1002/ange.201708273

Link to VoR: http://dx.doi.org/10.1002/anie.201708273
http://dx.doi.org/10.1002/ange.201708273

WILEY-VCH

10.1002/anie.201708273

WILEY-VCH

The First "Zero-Length" Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis

Christoph Hage[#], Claudio Iacobucci[#], Anne Rehkamp, Christian Arlt, and Andrea Sinz*

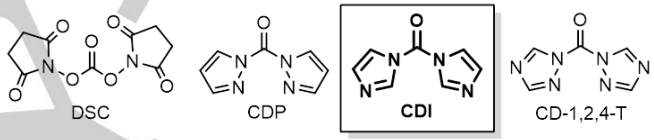
Abstract: Combining the properties of a "zero-length" cross-linker with cleavability by tandem mass spectrometry (MS/MS) poses great advantages for protein structure analysis using the cross-linking/MS approach. These include a reliable, automated data analysis and the possibility to obtain short-distance information of protein 3D-structures. We introduce *1,1'*-carbonyldiimidazole (CDI) as an easy-to-use and commercially available, low-cost reagent that ideally fulfils these features. CDI bridges primary amines and hydroxyl groups in proteins with the lowest possible spacer length of one carbonyl unit (~2.6 Å). The cross-linking reaction can be conducted at physiological conditions in the pH range between 7.2 and 8. Urea and carbamate cross-linked products are cleaved upon collisional activation during MS/MS experiments generating characteristic product ions, greatly improving the unambiguous identification of cross-links. Our innovative analytical concept is exemplified and applied for bovine serum albumin (BSA), wild-type tumor suppressor p53, an intrinsically disordered protein, and retinal guanylyl cyclase activating protein-2 (GCAP-2).

Chemical cross-linking/mass spectrometry (MS) has evolved as a powerful tool for 3D-structural analysis of proteins and protein complexes.[1] The cross-linker has a defined length and acts as a "molecular ruler", imposing distance constraints between connected amino acids. After the cross-linking reaction, enzymatic digestion of the covalently connected protein(s) yields complex peptide mixtures that are notoriously difficult to handle. Accurate and automated identification of cross-links is currently best achieved by using MS-cleavable cross-linkers.[2] These novel linkers, such as disuccinimidyl sulfoxide (DSSO)[2c] and disuccinimidyl dibutyric urea (DSBU or BuUrBu)[2f], are commercially available and drastically diminish potential false-positive identification rates of cross-links.

The spacer lengths of the MS-cleavable cross-linkers are 10.1 Å for DSSO and 12.5 Å for DSBU, allowing bridging side chains of amino acids that are spatially apart from each other. DSSO and DSBU are highly valuable for thoroughly mapping large protein assemblies, however, they do not ensure to derive geometrical constraints for higher resolution computational modeling due to their relatively long spacers.

Along the lines of extending our analytical concept of designing urea-based MS-cleavable cross-linkers, we aimed at shortening the spacer of DSBU to obtain more valuable geometrical constraints for protein modeling. In this work, we describe *1,1'*-carbonyldiimidazole (CDI) as the first cross-linker merging urea-based MS-cleavability with virtually "zero-length" (2.6 Å) spacer length.

Phosgene and its derivatives[3] are well-known reagents for the preparation of urea compounds under anhydrous conditions. Interestingly, Padiya *et al.*[4] described imidazole carbonylation of amines in water for preparing urea and carbamate compounds. Inspired by this finding, we envisioned the use of phosgene derivatives as amine- and hydroxyl-reactive, ultra-short cross-linkers. We screened the reactivities of four reagents (Scheme 1) towards three different proteins at various temperatures and pH values. This systematic study allowed us to identify CDI as effective protein cross-linker (Table S1, SI). The inefficacy of the three other reagents can be explained by either their low reactivity or fast hydrolysis rate. This is in agreement with the hydrolysis kinetics reported by Staab[5] who suggested the following reactivity order of CDP < CDI < CD-1,2,4-T.



Scheme 1. Structure of reagents (disuccinimidyl carbonate (DSC), *1,1'*-carbonyldiimidazole (CDI), *1,1'*-carbonyl-di-1,2,4-triazole (CD-1,2,4-T), *1,1'*-carbonyldipyrazole (CDP)) that were tested as urea-based, ultra-short protein cross-linkers. CDI has been identified herein as the most effective reagent.

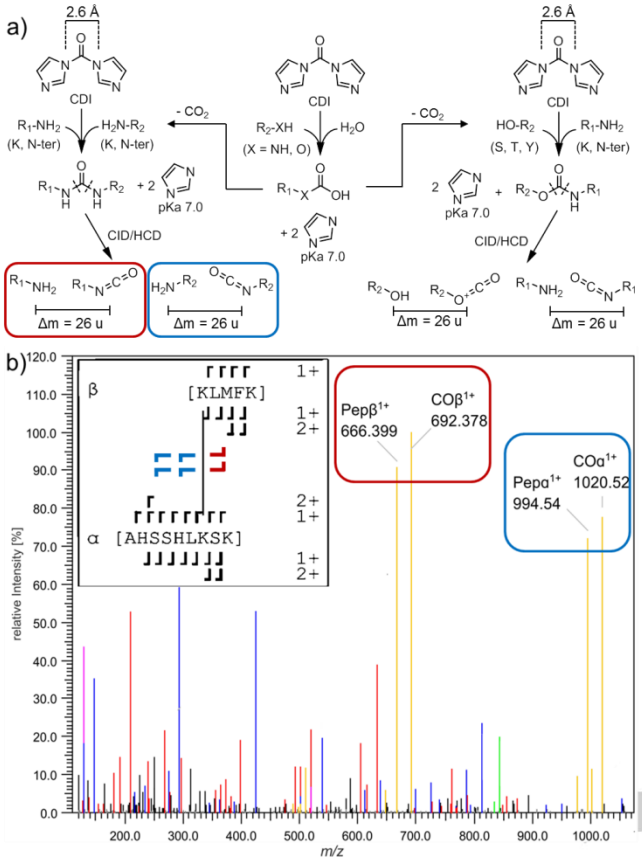
Reactivity. CDI belongs to an extraordinary class of electrophilic reactive amides, termed azolides, where the amide nitrogen is part of a quasi-aromatic five-membered heterocycle containing at least two nitrogens, i.e. azoles.[5] The amide nitrogen of azolides is 'pyrrole-like' and its lone electron pair contributes to the aromatic sextet of the azole. This significantly reduces the partial double bond character of the amide bond as well as the electron density of the nitrogen atom. This polarizes the exocyclic bond even further, increasing the electrophilicity of the central carbonyl. The reactivity of azolides towards nucleophiles is comparable to those of acyl halides and anhydrides, but they benefit from a tunable reactivity and hydrolysis rate, depending on the number and position of nitrogen atoms in the ring.[5] Among the azolides evaluated in this study (CDP, CDI, and CD-1,2,4-T; Scheme 1; Table S1, SI), CDI emerged as efficient cross-linker that is applicable in an aqueous environment at physiological pH conditions. CDI is inexpensive (~ 0.30 €/g) and cross-links proteins in the pH range between 7.2 and 8.0 at temperatures higher than 10°C.

CDI has been found to react with primary amines in lysine side chains yielding urea products. Interestingly, CDI has also been

* Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Martin Luther University Halle-Wittenberg Wolfgang-Langenbeck-Str. 4, D-06120 Halle/Saale (Germany), E-mail: andrea.sinz@pharmazie.uni-halle.de

# Both authors contributed equally to this paper.

SI for this article is given via a link at the end of the document.



Scheme 2. a) Reactivity of CDI with amines (lysines, *N*-termini) and hydroxyls (serines, threonines, tyrosines) in proteins. b) Diagnostic fragmentation pattern of the urea moiety upon collisional activation in tandem MS experiments. The two characteristic doublets are highlighted in the product ion mass spectrum of a doubly charged ion at *m/z* 843.4611 (shown in green). The tandem mass spectrum has been automatically assigned using MeroX[6] and identified as a CDI-cross-linked product between Lys-370 and Lys-382 (^{364}AHSSHL**K**SK^{372} – ^{382}**K**LMFK^{386}) in p53. The characteristic doublet signals caused by cleavage of the CDI cross-linker are shown in yellow, b-and y-type fragment ions of the peptide backbone are shown in blue and red.

found to connect primary amines in lysines with hydroxyl groups of serines, threonines, and tyrosine sides giving carbamates (Scheme 2a). Overall, we identified ~ 43% Lys-Lys and ~ 57% Lys-Ser/Thr/Tyr cross-links. The large percentage of cross-links involving hydroxyl groups can be explained by a higher reactivity of CDI. Also, the carbamates created with CDI are more stable in aqueous solution compared to the esters formed with hydroxyl groups using *N*-hydroxysuccinimide (NHS)-based cross-linkers.[7]

Remarkably, peptides modified with a partially hydrolyzed cross-linker (type-0 or so-called "dead-end" cross-links)[8] were not identified as the created carbamic acid and carbonate compounds are unstable. Carbamic acids decarboxylate and release the free amine or alcohol, which can further react with another CDI molecule. This presents a distinct advantage of CDI compared to conventional amine-reactive NHS-based cross-linkers, which undergo irreversible hydrolysis to form stable carboxylic acid derivatives.[9] The absence of peptides that are modified by a partially hydrolyzed cross-linker eliminates one of the most common sources of cross-links misassignments.[9] In fact, cross-links composed of consecutive amino acid sequences are isobaric to peptides involving identical amino acid sequences that are modified by a partially hydrolyzed cross-linker (Figure S1, SI).[9]

MS-cleavability. Upon collisional activation, CDI cross-linked products follow the well-known fragmentation pathway of the urea-based DSBU linker[2f]. In detail, two characteristic 26-u doublets are visible in the product ion mass spectrum, which allows distinguishing between different cross-link types. In case of the CDI cross-linker, one of the two diagnostic product ion signals of each doublet corresponds to the unmodified peptide (Scheme 2). The second one depends on the nature of cross-linked amino acid. It can either correspond to the peptide decorated with i) an isocyanate group, in case of reaction with an amine group of lysine or the *N*-terminus; ii) an oxonium ion, in case of reaction with a hydroxyl group of serine, threonine, and tyrosine (Scheme 2a). Exemplary product ion spectra of CDI-cross-links are presented in Scheme 2 and in the Supporting Information (SI).

In case of cross-linking a primary amine and an alcohol, the resulting product ion spectrum does not always show the complete doublet-of-doublet pattern. The protonation of the carbamate moiety, which triggers its cleavage in the gas-phase, is influenced by the lower gas-phase basicity of the ester oxygen compared to that of the amide nitrogen. Moreover, the unstable oxonium ion, created by cleavage of the amide bond of the carbamate, might suffer further CO_2 loss. Nevertheless, the partial lack of characteristic product ions does not hamper a safe and potentially automated cross-link identification. Recently, Liu *et al.* reported an optimized strategy for the automated identification of DSSO cross-linked products using only one of the four diagnostic fragments.[2b]

BSA cross-linking. We employed BSA as model protein to evaluate the reactivity of DSC, CDI, CD-1,2,4-T, and CDP. All four compounds were reacted for 30 minutes at two pH values (7.5 and 8.0), three temperatures (0, 10, 20°C), and at equimolar concentration as well as at 10-, 20-, and 50-fold molar excess compared to BSA (experimental details are given in the SI). CDI emerged as a powerful reagent for protein cross-linking at temperatures above 10°C and at both pH values (Figures S2 and S3, SI). Applying our established beam-type collision induced dissociation (HCD)-tandem MS strategy for analyzing cross-linked products in combination with the MeroX software[6] for data analysis, we identified 26 unique cross-links for BSA. The number of cross-link identifications slightly increased with temperature and pH (Figure S2, SI). Remarkably, the high reproducibility was striking for all reaction conditions examined in this study (Figure S3, SI). To verify the validity of our data, cross-links were mapped onto the published structure of BSA (PDB ID: 3v03, resolution: 2.7 Å) to determine the Cα-Cα distances of these residues that were connected within the BSA monomer (Figure S4, SI). The Cα-Cα distance for elongated lysine side chains is ~13.4 Å plus ~2.6 Å for the carbonyl group of CDI, which results in a distance of ~16 Å to be bridged by CDI. The average distance measured for BSA was 12.5 Å. A maximum distance of 17.8 Å was found in our cross-linking

experiments, which can be explained by the flexibility of BSA and the X-ray structure's resolution. The distribution of Cα-Cα distances is provided in Table S2, SI. These results suggest that the cross-linking conditions do not disturb the conformation of BSA. The low molar excess required for CDI cross-linking has a crucial role in minimizing eventually detrimental effects on the protein structure by excessive cross-linking. These results encouraged us to further apply this reagent to structural investigations of the tumor suppressor protein p53, an intrinsically disordered protein (IDP), and GCAP-2.

p53 cross-linking. The short spacer length of CDI may be particularly beneficial for studying 3D-structures of IDPs. These proteins, which comprise ~25% of mammalian proteomes, are mainly involved in cell signaling and are characterized by a lack of defined 3D-structures.[10] The gold standard for the structural analysis of IDPs is currently NMR spectroscopy.[11] We employed CDI to investigate the elusive structure of the full-length, wild-type human tumor suppressor p53. Known as "guardian of the genome", human p53 consists of 393 amino acids that are organized as biologically active homotetramer, for which different structural concepts have been proposed. P53 cross-linking has been performed at pH 7.2 at 10°C in the absence of DNA yielding 23 unique cross-links (see SI). Interestingly, the cross-links were located not only in the *C*-terminal region, harboring the majority of p53's lysines, but also in the tetramerization domain and the DNA-binding domain (Figure S5, SI). The latter p53 domain has never been connected by any of the cross-linkers employed previously[14]. The identification of complementary cross-links in p53 is probably mainly attributed to an improved reactivity of CDI with hydroxyl groups compared to NHS ester cross-linkers[14]. This underlines the great potential of CDI to shed light onto the spatial arrangement of p53 and IDPs in general.

Unique CDI-cross-links were mapped onto two modeled structures of the p53 tetramer, as proposed by Tidow *et al.*[12] and by Okorokov *et al.*[13], to assess whether they are consistent with one of p53 models. As mentioned above, the maximum Cα-Cα distance to be bridged by CDI is ~16 Å, which however might be considered as too stringent in this case.

The great majority of cross-links was found to be consistent with the intramolecular Cα-Cα distances measured in the model proposed by Okorokov *et al.*[13] These cross-links fall within the range of 6.6 to 30 Å with an average of 19.3 Å and those exceeding the maximum bridging distance of ~16 Å are located in flexible loop regions[13]. In contrast, the SAXS-based model by Tidow *et al.*[12] does not agree with our cross-linking data. Here, average of Cα-Cα distances is 42.4 Å for intramolecular 65.5 Å for intermolecular cross-links (Figure S6, SI).

Therefore, our CDI cross-linking results strongly support the structural organization of the p53 tetramer as proposed by Okorokov *et al.*[13], which presents a more compact arrangement than the elongated cross-shaped structure introduced by Tidow *et al*.[12]. This finding is also in good agreement with recent p53 cross-linking experiments, in which DSBU was employed.[14] Compared to DSBU, CDI yields ~33% shorter geometrical constraints and suggests a more compact structure of p53 as perceived in the p53 model proposed by Tidow *et al*.[12].

Interestingly, none of the intermolecular cross-links identified with DSBU[14] have been observed for CDI. We assume that the relevant residues are sufficiently far apart to prevent CDI cross-linking. This again testifies the valuable complementarity of the MS-cleavable cross-linkers CDI and DSBU for structure analysis of a large variety of proteins.

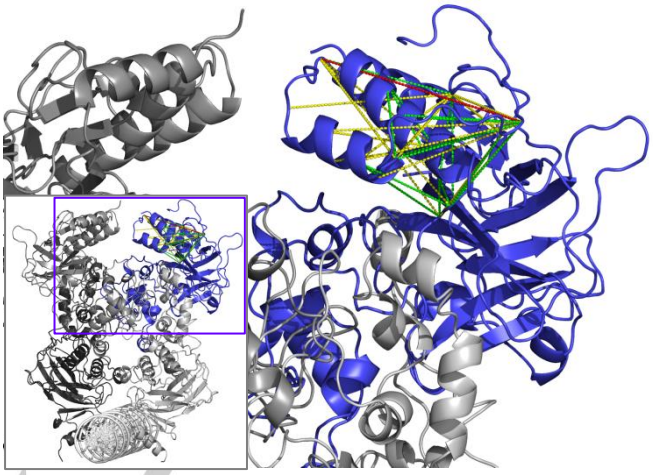


Figure 1. Cα–Cα distances bridged by the CDI cross-linker are mapped into p53 tetrameric model structure by *Okorokov et al.*[13]. Cα–Cα distances of cross-linked residues are indicated as colored lines; green: distances <17 Å; yellow: distances <27 Å; red: distances >27 Å.

Guanylyl cyclase-activating protein-2 (GCAP-2) cross-linking

GCAP-2 is a Ca^{2+}-binding protein that belongs to the neuronal calcium sensor protein family.[15] GCAPs regulate guanylyl cyclases in the retina cells and play a key role for light adaptation.[16] *N*-terminally myristoylated, Ca^{2+}-bound bovine GCAP-2 was reacted with CDI yielding 19 unique cross-links (see SI). They were compared with previously obtained BS^2G cross-links of GCAP-2[17], showing 8 identical cross-links (six intramolecular and two intermolecular ones). Among the cross-links of GCAP-2 monomer, three of them could be mapped into the NMR structure of Ca^{2+}-bound GCAP-2 (PDB ID: 1JBA, resolution: 3.5 - 5.2 Å)[18]. The Cα-Cα distances measured were in the range between 14.6 to 17.4 Å, which are in perfect agreement with the distance CDI can bridge. The remaining three cross-links, identified with both CDI and BS^2G, involve K-200 (or S-201) located in the *C*-terminal domain of GCAP-2 (S-191 to F-204), which is structurally not resolved in the NMR structure. The four amino acids involved, K-30, K-50, K-96, K-126 (or K-128), cover a triangular surface on GCAP-2 (Figure 2). Interestingly, this surface is located on the opposite side of the three Ca^{2+}-binding EF-hand motifs. The most *C*-terminal amino acid resolved in the NMR structure is P-190 that is located above this triangular surface. The cross-links between S-37 (or T-39) and K-200 as well as between K-50 and K-200 (or S-201) can be explained by a high flexibility of GCAP-2's *C*-terminus. However, the *C*-terminus of GCAP-2 was not found to be cross-linked with amino acids in the EF-hand motifs of GCAP-2 that are located on the opposite site of the molecule.

Among the remaining 9 cross-links that were exclusively identified with CDI only two are in agreement with the structure

of Ca^{2+}-bound GCAP-2. The remaining seven cross-links exceed the CDI distance limit within the GCAP-2 monomer, but they can be explained by the GCAP2 dimer.

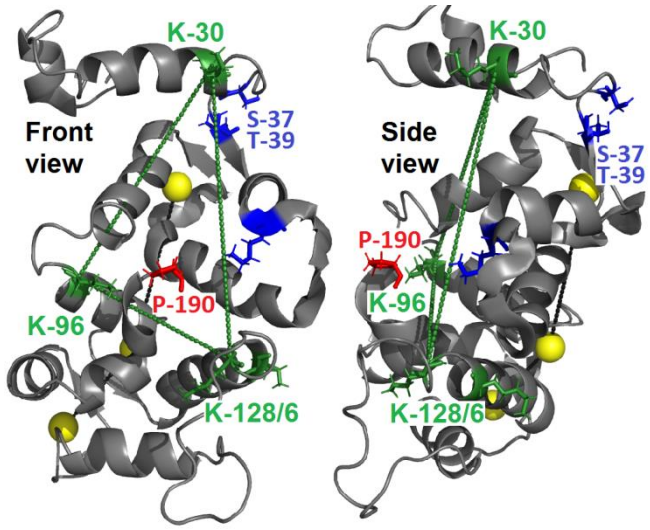


Figure 2. Front and side views of GCAP-2. The flexible *C*-terminal region of GCAP-2 (S-191 to F-204) is not resolved in the NMR structure (PDB ID: 1JBA)[18]. The most *C*-terminal residue resolved in the NMR structure, P-190 (red), resides above a triangular surface (green dashed lines) that is formed by the lysines K-30, K-96, K-126/128 (green) cross-linked with K-200 (or S201). K50, S-37 (or T-39) (blue), were cross-linked to K-200 (or S-201) only by CDI. The opposite site of GCAP-2 is defined by three Ca^{2+} ions (yellow spheres), shown as black dashed line.

Conclusion. Our ongoing search for novel cross-linking principles led to the identification of the azolide CDI cross-linker. CDI is the first "zero-length" cross-linker that is cleavable under MS/MS conditions and delivers characteristic product ions that facilitate the analysis of cross-linked products in protein conformational studies. The cross-linking reaction can be conducted at near-physiological pH conditions between 7.2 and 8. The short spacer length of CDI yields complementary distance information compared to the so far mainly used NHS-based cross-linkers. As CDI possesses similar reactivites towards amine and hydroxyl groups in proteins, it allows targeting additional regions in proteins that have so far been not accessible to cross-linking with NHS esters. Also, sample complexity is reduced due to the absence of "dead-end" (type 0) cross-links. Other advantages of CDI are its low costs and its straightforward application that will make the cross-linking/MS approach available to a large number of laboratories having access to mass spectrometers with MS/MS capabilities.

Acknowledgements

The authors would like to thank X. Wang for her contribution in data analysis of BSA cross-links and D. Tänzler for GCAP-2 expression and purification. Dr. C. Ihling is acknowledged for his valuable advice regarding LC/MS/MS analyses. A.S. acknowledges financial support by the DFG (project Si 867/15-2). C.I. is funded by the Alexander von Humboldt Foundation.

Keywords: *1,1'*-Carbonyldiimidazole (CDI) • cross-linking/mass spectrometry (MS) • MS-cleavable cross-linker • p53 • GCAP- • protein structure.

[1] a) A. Sinz, *Mass Spectrom. Rev.* **2006**, *25*, 663-682; b) J. Rappsilber, *J Struct Biol* **2011**, *173*, 530-540; c) A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, *Mol. Cell. Proteomics* **2010**, *9*, 1634-1649; d) C. Iacobucci, S. Reale, F. De Angelis, *Chembiochem* **2013**, *14*, 181-183.

[2] a) F. Liu, D. T. S. Rijkers, H. Post, A. J. R. Heck, *Nat. Methods* **2015**, *12*, 1179-1184; b) F. Liu, P. Lössl, R. Scheltema, R. Viner, A. J. R. Heck, *Nat. Commun* **2017**, *8*; c) A. H. Kao, C. L. Chiu, D. Vellucci, Y. Y. Yang, V. R. Patel, S. H. Guan, A. Randall, P. Baldi, S. D. Rychnovsky, L. Huang, *Mol. Cell. Proteomics* **2011**, *10*; d) C. Arlt, M. Gotze, C. H. Ihling, C. Hage, M. Schafer, A. Sinz, *Anal. Chem.* **2016**, *88*, 7930-7937; e) C. Iacobucci, C. Hage, M. Schafer, A. Sinz, *J. Am. Soc. Mass Spectr.* **2017**; f) M. Q. Müller, F. Dreiocker, C. H. Ihling, M. Schäfer, A. Sinz, *Anal. Chem.* **2010**, *82*, 6958-6968.

[3] F. Bigi, R. Maggi, G. Sartori, *Green Chem.* **2000**, *2*, 140-148.

[4] K. J. Padiya, S. Gavade, B. Kardile, M. Tiwari, S. Bajare, M. Mane, V. Gaware, S. Varghese, D. Harel, S. Kurhade, *Org. Lett.* **2012**, *14*, 2814-2817.

[5] H. A. Staab, *Angew. Chem. Int. Ed.* **1962**, *1*, 351, *Angew. Chem.* **1962**, *74*, 407.

[6] M. Götze, J. Pettelkau, R. Fritzsche, C. H. Ihling, M. Schäfer, A. Sinz, *J. Am. Soc. Mass Spectr.* **2015**, *26*, 83-97.

[7] C. L. Swaim, J. B. Smith, D. L. Smith, *J. Am. Soc. Mass Spectr.* **2004**, *15*, 736-749.

[8] B. Schilling, R. H. Row, B. W. Gibson, X. Guo, M. M. Young, *J. Am. Soc. Mass Spectr.* **2003**, *14*, 834-850.

[9] C. Iacobucci, A. Sinz, *Anal. Chem.* **2017**, *89*, 7832-7835.

[10] a) A. C. Joerger, A. R. Fersht, *Annu. Rev Biochem.* **2016**, *85*, 375-404; b) A. C. Joerger, A. R. Fersht, *Annu. Rev Biochem.* **2008**, *77*, 557-582.

[11] J. T. Nielsen, F. A. Mulder, *Front. Mol. Biosci.* **2016**, *3*, 4.

[12] H. Tidow, R. Melero, E. Mylonas, S. M. Freund, J. G. Grossmann, J. M. Carazo, D. I. Svergun, M. Valle, A. R. Fersht, *P. Natl Acad. Sci. USA* **2007**, *104*, 12324-12329.

[13] a) A. L. Okorokov, M. B. Sherman, C. Plisson, V. Grinkevich, K. Sigmundsson, G. Selivanova, J. Milner, E. V. Orlova, *The EMBO J.* **2006**, *25*, 5191-5200; b) R. Aramayo, M. B. Sherman, K. Brownless, R. Lurz, A. L. Okorokov, E. V. Orlova, *Nucleic Acids Res.* **2011**, *39*, 8960-8971.

[14] a) C. Arlt, V. Flegler, C. H. Ihling, M. Schäfer, I. Thondorf, A. Sinz, *Angew. Chem. Int. Ed.* **2017**, *56*, 275-279; *Angew. Chem.* **2017**, *129*, 281-285; b) C. Arlt, C. H. Ihling, A. Sinz, *Proteomics* **2015**, *15*, 2746-2755.

[15] W. A. Gorczyca, M. P. Graykeller, P. B. Detwiler, K. Palczewski, *P. Natl Acad. Sci. USA* **1994**, *91*, 4014-4018.

[16] K. W. Koch, D. Dell'Orco, *ACS Chem. Neurosci.* **2013**, *4*, 909-917.

[17] a) J. Pettelkau, I. Thondorf, S. Theisgen, H. Lilie, T. Schröder, C. Arlt, C. H. Ihling, A. Sinz, *J. Am. Soc. Mass Spectr.* **2013**, *24*, 1969-1979; b) J. Pettelkau, T. Schröder, C. H. Ihling, B. E. S. Olausson, K. Kölbel, C. Lange, A. Sinz, *Biochemistry* **2012**, *51*, 4932-4949.

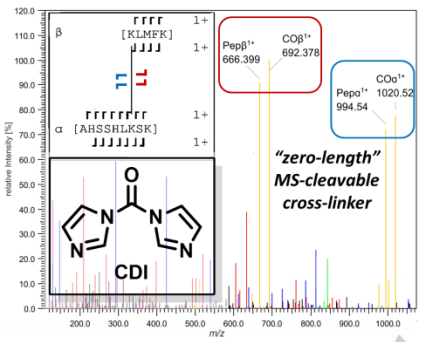
[18] J. B. Ames, A. M. Dizhoor, M. Ikura, K. Palczewski, L. Stryer, *J. Biol. Chem.* **1999**, *274*, 19329-19337.

10.1002/anie.201708273

WILEY-VCH

COMMUNICATION

1,1'-carbonyldiimidazole (CDI) is the shortest possible (~2.6 Å) urea-based, MS cleavable cross-linker. It reacts with amine and hydroxyl groups in proteins at physiological pH. CDI has been successfully applied to cross-link BSA, the tumor suppressor p53, and GCAP-2. It allows deriving valuable short-distance constraints in proteins, which proved especially valuable for probing intrinsically disordered proteins (IDPs), such as p53.



*Christoph Hage[#], Claudio Iacobucci[#], Anne Rehkamp, Christian Arlt, and Andrea Sinz**

Page No. – Page No.