Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Journal Name

RSCPublishing

COMMUNICATION

Twisted amide electrophiles enable cyclic peptide sequencing

Cite this: DOI: 10.1039/xoxxoooooox

Serge Zaretsky,[a] Vishal Rai,[a] Gerald Gish,[b] Matthew W. Forbes,[a] Michael Kofler,[b] Joy C. Y. Yu,[a] Tony Pawson,[b] and Andrei K. Yudin[a]*

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

There is an ever-increasing interest in synthetic methods that not only enable peptide macrocyclization, but also facilitate downstream application of the synthesized molecules. We have found that aziridine amides are stereoelectronically attenuated in a macrocyclic environment such that non-specific interactions with biological nucleophiles are reduced or even shut down. The electrophilic reactivity, revealed at high pH, enables peptide sequencing by mass spectrometry, which will further broaden the utility of aziridine amide-containing libraries of macrocycles.

Macrocycles continue to attract attention on behalf of the chemistry community.[1] Their unique structural features have found many uses in areas such as medicinal chemistry,[2] coordination chemistry,[3] and nanotechnology.[4] Cyclic peptides, as a subset of macrocycles, are noteworthy for their biological applications,[5] specifically in the field of protein-protein interactions.[6] Our group has been interested in exploring the conformational consequences of placing strained amides within the structures of cyclic peptides. Herein we demonstrate the surprising compatibility of aziridine amides with biological nucleophiles, and also showcase the utility of preserving the latent electrophile in a cyclic peptide sequencing application.

Torsional strain provides an opportunity to stereoelectronically "challenge" one of nature's most ubiquitous structural units – the amide bond.[7] The electrophilic reactivity of the amide carbon is modulated through delocalization of the nitrogen lone pair of electrons, but in twisted amides this conjugation is significantly diminished.[8] Seminal studies in the area of twisted amides have focused on using anti-Bredt tricyclic systems that enforce the nitrogen lone pair out of conjugation with the acyl substituent.[9] Kirby and co-workers have shown that the nitrogen center in 1-aza-2-adaman-tanone displays nucleophilic amine reactivity, while the carbonyl group acts as a ketone-like electrophile.[10] A study of twisted lactams by the Aubé group suggested that a minimum 50° distortion of the amide was required to observe nucleophilic amine reactivity.[11] By resorting to the Schmidt-Aubé reaction, Stoltz and co-workers have prepared the long-sought 2-quinuclidone as a HBF_4 salt.[12] While these examples highlight the unusual properties of twisted amides, they do not suggest obvious applications.

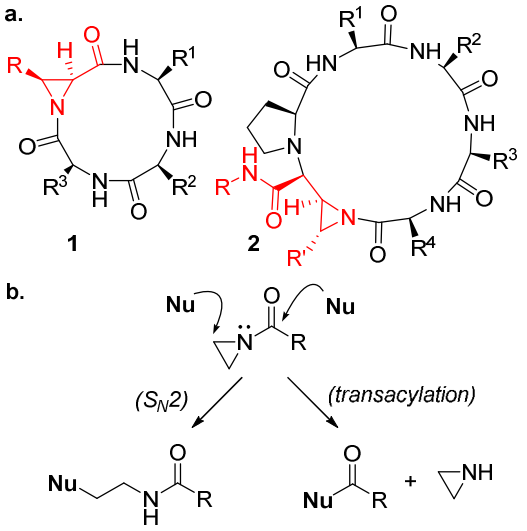

Figure 1. (a) Aziridine amide-bearing cyclic peptides made by peptide cyclization with an Azy residue (**1**) or by aziridine aldehyde-mediated macrocyclization (**2**); (b) Aziridine amide bis-electrophile reactivity by S_N2 or transacylation.

We have developed several methods for generating macrocycles with embedded aziridine amides (Figure 1a)[13] and demonstrated downstream reactivity of these compounds with numerous nucleophiles.[14] Aziridine amides are *ambident* electrophiles, meaning that they can partake in both S_N2 and transacylation reactions (Figure 1b).[15] This dual electrophilicity is caused by the strained three-membered aziridine ring and the electrophilic carbonyl group, which exhibits diminished delocalization of the nitrogen lone pair of electrons compared to regular amides.[16] As reactive intermediates, aziridine amides have been primarily used in S_N2 reactions with soft nucleophiles.[17] Meanwhile, hard

nucleophiles are known to trigger the transacylation reaction, a process that has gained comparatively little attention.[18]

We now report that stereoelectronic attenuation of the aziridine amide can prevent non-specific reactivity with biological nucleophiles. Additionally, we report that preserving the electrophilic carbonyl group of the aziridine amide enables cyclic peptide sequencing by chemoselective linearization and tandem mass spectrometry.

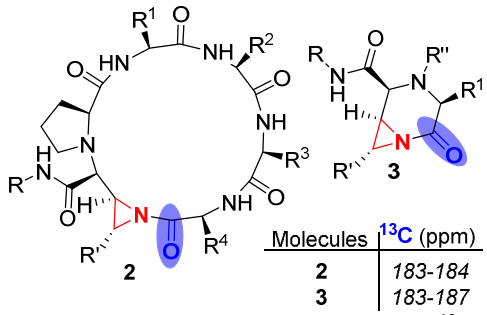

Figure 2. Comparison of ketone-like character in ^{13}C of larger cycles (**2**) vs. 6-membered rings (**3**). **2a** & **2b**: R = ^{t}Bu, R^1 = H, R^2 = $CH_2CH(CH_3)_2$, R^3 = H, R^4 = $CH_2C_6H_5$, R' = H (**2a**); R' = $CH_2CH(CH_3)_2$ (**2b**).

In piperazinones (**3**), the nitrogen lone pair cannot effectively conjugate with the carbonyl π* orbital, which is evident in the ^{13}C chemical shift of 183-187 ppm ($CDCl_3$). Within larger rings, we noted a slightly more upfield chemical shift of the aziridine amide carbonyl group (Figure 2). Within 18-membered rings, the aziridine carbonyl shift appears at 183-184 ppm ($CDCl_3$), which is closer to the chemical shift of aziridine amides in linear systems (181-183 ppm, $CDCl_3$).[19] While the differences are small, the data still suggests a tendency for 6-membered rings to posses more ketone-like character in the aziridine amide. Our hypothesis was that larger rings would engender aziridine amides with greater stability towards hydrolysis and reduced non-specific reactivity through the transacylation pathway.[20] At the same time, we were interested in preventing non-specific S_N2 reactivity, which would be promoted by the increased N-to-C conjugation.

As a testing ground, we decided to evaluate a series of probes **3a**, **2a**, and **2b** (Figure 2) in a biological setting. **3a** was made from Pro and cyclic peptides **2a** and **2b** were prepared from H-Pro-Gly-Leu-Gly-Phe-OH and the respective aziridine aldehyde reagents using our previously reported methods.[13b] When **2a** and **2b** were evaluated for their reactivity with glutathione (GSH), only the unsubstituted aziridine analog **2a** was found to react (Figure 3a). The thiol attack at the aziridine was confirmed by collision-induced dissociation/tandem mass spectrometry (CID/MS^2). Owing to its unhindered aziridine and non-attenuated acyl reactivity, probe **3a** reacted with GSH by both the S_N2 and transamidation pathways.

We next investigated any reactivity against a broader biological nucleophile, a protein. For our goals, we sought a protein that: a) contained a number of surface-exposed nucleophilic residues and b) was capable of enzymatic activity. The Fes SH2-kinase was chosen as a model.[21] The catalytic SH2-kinase domain of the protein tyrosine kinase Fes is a complex nucleophile containing eight surface cysteine residues with differentiated reactivity,[22] along with a variety of other nucleophilic amino acid side chains. Cyclic aziridine amides **3a** and **2a** were tested for their ability to covalently label Fes SH2-kinase.[23] As a positive control, we evaluated the reactivity of Fes SH2-kinase against iodoacetamide, a cysteine-selective electrophile.[24] The thiol groups of the surface-exposed cysteine residues in Fes were completely converted into thioacetamides (up to eight alkylations, see Supporting Information) following treatment with 9 mM iodoacetamide. Compounds **3a** and **2a** were tested under similar conditions. Up to four adducts of **3a** were observed in 2 hours, but no reactivity was detected with **2a** (Figure 3b). To deduce the type of electrophilic reactivity of **3a**, the Fes SH2-kinase/**3a** adduct was peralkylated with iodoacetamide. Eight thioacetamide alkylations were observed, ruling out thiol attack at the aziridine. The formation of adducts with **3a** can be explained by transamidation with lysine residues found on the protein surface. This is in good agreement with empirical evidence of carbonyl electrophilicity in smaller aziridine-containing ring systems (Figure 2).[13b] The lack of reactivity with **2a** suggests that, while the aziridine is susceptible to nucleophilic attack with small nucleophiles (eg. GSH), non-specific reactivity on the protein surface is not favourable.

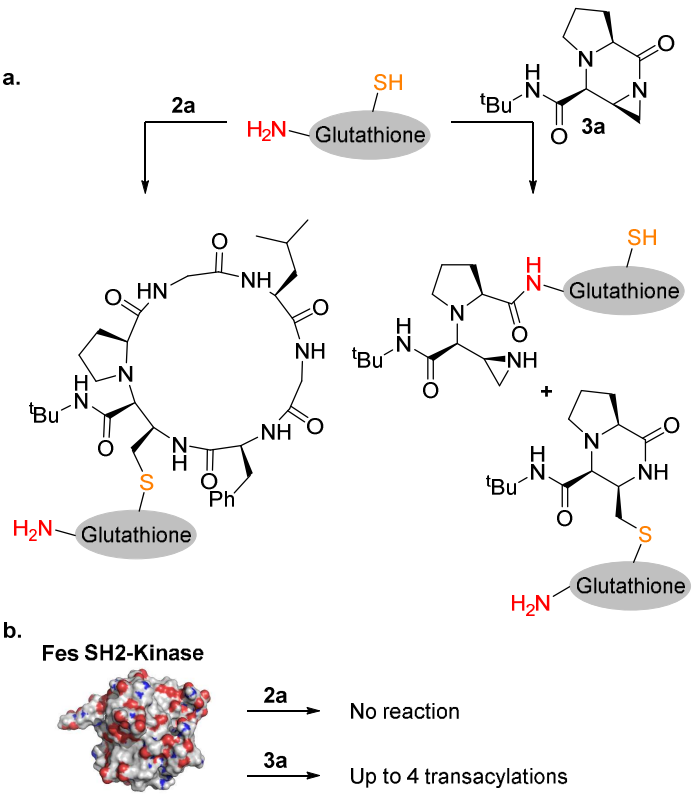

Figure 3. (a) Increasing amidicity of the strained aziridine amide bond in **2a** prevents the transacylation pathway seen for **3a** with GSH; (b) Exposing FES SH2-kinase to cyclic aziridine amides leads to no reactivity with **2a**, while **3a** reacts with the protein in a cysteine-independent manner.

With biological compatibility established for the 18-memberd ring macrocycles, we sought to further explore the "hard" carbon center of the acyl aziridine. When we subjected **2b** to basic aqueous media (0.8 M NaOH), clean and site-selective hydrolysis of the aziridine amide took place to yield compound **4** (Figure 4). This hydrolysis product was composed of the linear peptide segment attached to the cyclization-linker at the N-terminus. Site-selective linearization has been previously applied to cyclic peptoids,[25] but for cyclized peptides the technique is limited to ring scission by cyanogen bromide,[26] which is plagued by a number of side reactions.[27]

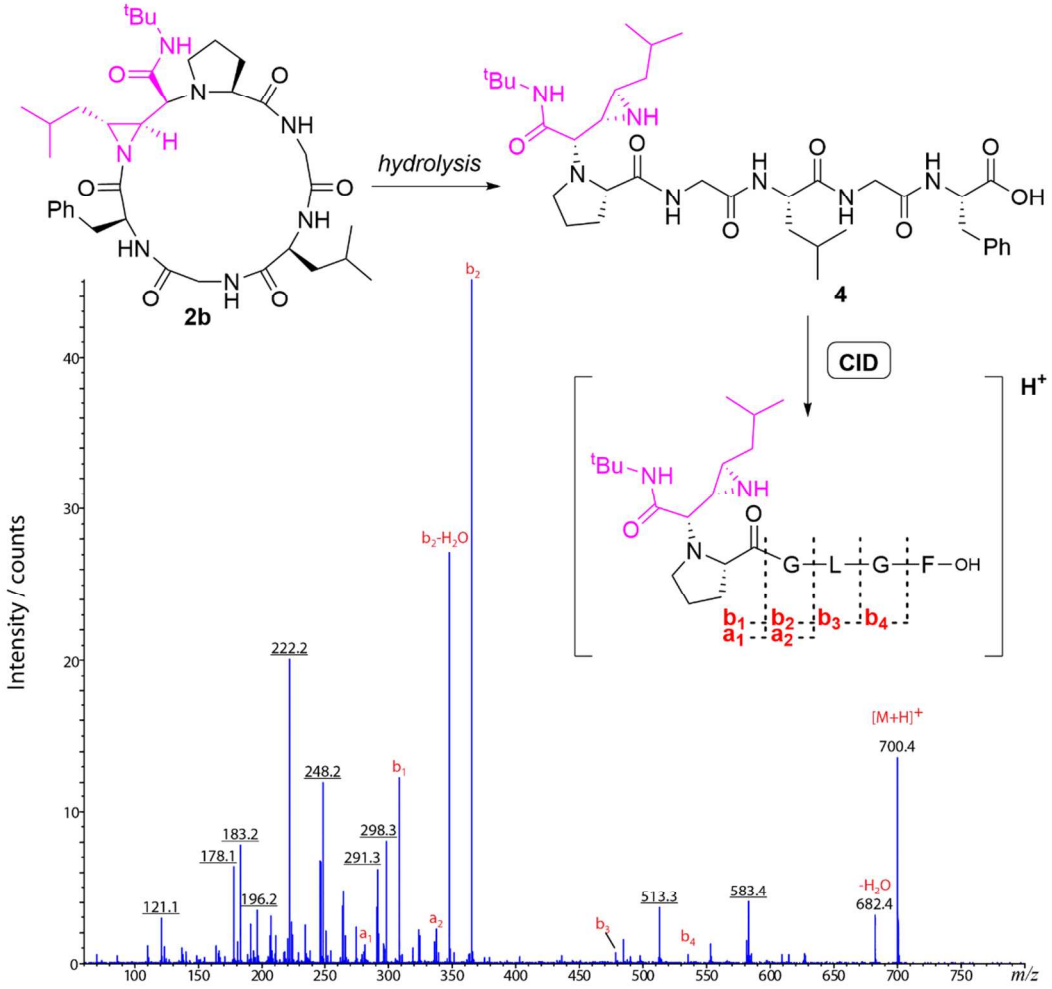

Figure 4. Base-mediated hydrolysis of **2b** to linearized peptide **4**. **4** is fragmented by CID/MS^2 to prominent *a*- and *b*- ions. Underlined ions correspond to characteristic fragmentation ions of the linker.

Conversely, regiospecific linearization at the aziridine amide is enabling because it allows for sequencing of macrocyclic peptides using CID/MS^2. Compound **4** fragmented readily with CID/MS^2 to characteristic *a*- and *b*- ions (Figure 3).[28] BioAnalyst[29] was used to annotate the spectra for sequence information using the most common 20 natural amino acids and a modified proline residue incorporating the cyclization linker. The highest scoring sequence tags corresponded to partial internal sequences and the highest scoring full sequence was the correct peptide sequence. The ability to site-specifically cleave and sequence aziridine-containing macrocycles is significant in that it allows for post-assay identification of a peptide. Hydrolysis with ^{18}O-labelled water may provide further assistance in deciphering the *y*-ions.[30]

The selectively linearized macrocyclic peptides do not lead to scrambled sequences and the cyclization tag serves as a reference point for downstream peptide sequencing. In contrast, conventional cyclic peptides can fragment at multiple positions, complicating interpretation of the CID/MS^2 spectra.[31] The cyclic topology also rules out Edman degradation for sequence elucidation and instead, elaborate protocols must be engineered to overcome these limitations.[32] Accordingly, the challenge of *de novo* peptide sequencing has limited biological screening techniques, such as affinity selection-mass spectrometry, to linear peptides and not their cyclic counterparts.[33] The aziridine amide enabled linearization technique should provide the tools needed to screen and sequence libraries of biologically stable cyclic peptide probes.[34]

In summary, we have evaluated the reactivity of aziridine amides embedded within peptidic rings and demonstrated that the 18-memberd ring macrocycles are stereoelectronically attenuated such that non-specific reactivity with biological nucleophiles such as glutathione or Fes SH2-kinase is reduced, or altogether absent. While aziridines are typically known for irreversible reactivity with biomolecules, our study opens the door to novel applications of this class of molecules. The key benefit of retaining a strained amide is the ability selectively hydrolyze it under harsh basic conditions. The chemoselective linearization of aziridine amide-containing macrocycles enabled routine sequencing by CID/MS^2 without the need for multistage mass spectrometry or complex algorithms.[35] Because the undesired S_N2 chemistry has been shut down, the macrocycles are amenable to site-specific modification (e.g. enzymes) followed by deconvolution with site-selective linearization and sequencing. These findings will lead to novel biological screening applications of cyclic peptides with embedded aziridine amides.

Funding for this work was provided by the Natural Sciences and Engineering Research Council (NSERC). S.Z. thanks the Ontario Graduate Scholarship (OGS) program for financial support.

Notes and references

[a] Dr. S. Zaretsky, Dr. V. Rai, Dr. M. W. Forbes, Ms. J. C. Y. Yu, Prof. Dr. A. K. Yudin. Davenport Research Laboratories, Department of

Chemistry, University of Toronto, 80 St. George St., Toronto, ON, M5S 3H6 (Canada) E-mail: ayudin@chem.utoronto.ca

[b] Dr. G. Gish, Dr. M. Kofler, Prof. Dr. T. Pawson. Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Av., Toronto, ON, M5G 1X5, (Canada).

†Electronic Supplementary Information (ESI) available: methods and CID/MS^2 analysis. See DOI: 10.1039/c000000x/

[1] A. K. Yudin, *Chem. Sci.*, 2015, **6**, 30–49.
[2] F. Giordanetto and J. Kihlberg, *J. Med. Chem.*, 2014, **57**, 278–95; (b) X. Yu and D. Sun, *Molecules*, 2013, **18**, 6230–6268; (c) J. Mallinson and I. Collins, *Future Med. Chem.*, 2012, **4**, 1409–1438; (d) E. M. Driggers, S. P. Hale, J. Lee, and N. K. Terrett, *Nat. Rev. Drug Discov.*, 2008, **7**, 608–624.
[3] (a) L. F. Lindoy, K.-M. Park, and S. S. Lee, *Chem. Soc. Rev.*, 2013, **42**, 1713–1727; (b) B. P. Burke and S. J. Archibald, *Annu. Rep. Prog. Chem., Sect. A: Inorg. Chem.*, 2013, **109**, 232-253.
[4] (a) R. J. Brea, C. Reiriz, and J. R. Granja, *Chem. Soc. Rev.*, 2010, **39**, 1448–1456; (b) R. Chapman, M. Danial, M. L. Koh, K. A Jolliffe, and S. Perrier, *Chem. Soc. Rev.*, 2012, **41**, 6023–6041.
[5] (a) S. E. Northfield, C. K. Wang, C. I. Schroeder, T. Durek, M.-W. Kan, J. E. Swedberg, and D. J. Craik, *Eur. J. Med. Chem.*, 2014, **77**, 248–257; (b) D. J. Craik and A. C. Conibear, *J. Org. Chem.*, 2011, **76**, 4805–4817;
[6] (a) L. N. Makley and J. E. Gestwicki, *Chem. Biol. Drug Des.*, 2013, **81**, 22–32; (b) E. Marsault and M. L. Peterson, *J. Med. Chem.*, 2011, **54**, 1961–2004; (c) M. Katsara, T. Tselios, S. Deraos, G. Deraos, M.-T. Matsoukas, E. Lazoura, J. Matsoukas, and V. Apostolopoulos, *Curr. Med. Chem.*, 2006, **13**, 2221–2232.
[7] (a) M. Szostak and J. Aubé, *J. Am. Chem. Soc.*, 2010, **132**, 2530-2531; (b) M. Szostak and J. Aubé, *Chem. Commun.*, 2009, 7122-7124; (c) M. Szostak, L. Yao and J. Aubé, *J. Org. Chem.*, 2009, **74**, 1869-1875; (d) M. Szostak and J. Aubé, *Org. Lett.*, 2009, **11**, 3878-3881; (e) M. Szostak and J. Aubé, *J. Am. Chem. Soc*., 2009, **131**, 13246-13247. (f) J. Clayden, W. J. Moran, *Angew. Chem. Int. Ed.*, 2006, **45**, 7118-7120.
[8] (a) M. Hutchby, C. E. Houlden, M. F. Haddow, S. N. G. Tyler, G. C. Lloyd-Jones and K. I. Booker-Milburn, *Angew. Chem. Int. Ed.*, 2012, **124**, 563-566; (b) M. Szostak, L. Yao and J. Aubé, *J. Am. Chem. Soc*., 2010, **132**, 2078-2084; (c) V. Somayaji and R. S. Brown, *J. Org. Chem.*, 1986, **51**, 2676-2686.
[9] (a) A. J. Kirby, I. V Komarov, K. Kowski and P. Rademacher, *J. Chem. Soc., Perkin Trans. 2*, 1999, **120**, 1313-1316; (b) A. J. Kirby, I. V Komarov and N. Feeder, *J. Am. Chem. Soc*., 1998, **120**, 7101-7102; (c) A. J. Kirby, I. V. Komarov, P. D. Wothers and N. Feeder, *Angew. Chem. Int. Ed.*, 1998, **37**, 785-786.
[10] A. J. Kirby, I. V. Komarov and N. Feeder, *J. Chem. Soc., Perkin Trans. 2*, 2001, 522-529.
[11] M. Szostak, L. Yao, V. W. Day, D. R. Powell and J. Aubé, *J. Am. Chem. Soc*., 2010, **132**, 8836-8837.
[12] (a) M. Szostak and J. Aubé, *Org. Biomol. Chem.*, 2011, **9**, 27-35; (b) M. Szostak, L. Yao and J. Aubé, *J. Org. Chem.*, *2010*, **75**, 1235-1243; (c) K. Tani and B. M. Stoltz, *Nature* 2006, **441**, 731-734.
[13] (a) C. J. White and A. K. Yudin, *Org. Lett.*, 2012, **14**, 2898–2901; (b) R. Hili, V. Rai, and A. K. Yudin, *J. Am. Chem. Soc.*, 2010, **132**, 2889–2891.
[14] (a) S. Zaretsky, J. Tan, J. L. Hickey, A. K. Yudin, in *Methods Mol. Biol. Pept. Libr.* (Ed.: R. Derda), Springer Science+Business Media, New York, **2015**, pp. 67–80; (b) S. Zaretsky, C. C. G. Scully, A. J. Lough, and A. K. Yudin, *Chem. - A Eur. J.*, 2013, **19**, 17668-17672; (c) A. Roxin, J. Chen, C. C. G. Scully, B. H. Rotstein, A. K. Yudin, and G. Zheng, *Bioconjug. Chem.*, 2012, **23**, 1387-1395; (d) B. H. Rotstein, R. Mourtada, S. O. Kelley, and A. K. Yudin, *Chem. - A Eur. J.*, 2011, **17**, 12257-12261.
[15] (a) H. Mayr, M. Breugst and A. R. Ofial, *Angew. Chem. Int. Ed.*, 2011, **50**, 6470-6505; (b) A. E. Rosamilia, F. Arico and P. Tundo, *J. Org. Chem.*, 2008, **73**, 1559-1562; (c) P. Tundo, L. Rossi and A. Loris, *J. Org. Chem.*, 2005, **70**, 2219-2224; (d) S. Hünig, *Angew. Chem. Int. Ed.*, 1964, **3**, 548-560.
[16] S. A. Glover and A. A. Rosser, *J. Org. Chem.*, 2012, **77**, 5492-5502.
[17] (a) X. E. Hu, *Tetrahedron*, 2004, **60**, 2701-2743; (b) H. Stamm, *J. Prakt. Chem.*, 1999, **341**, 319-331; (c) T. Schirmeister, *Arch. Pharm.*, 1996, **329**, 239-244; (d) T. Mall and H. Stamm, *J. Org. Chem.*, 1987, **52**, 4812-4814.
[18] (a) J. B. Sweeney, *Chem. Soc. Rev.*, 2002, **31**, 247-258; (b) B. Lygo, *Tetrahedron*, 1995, **51**, 12859-12868.
[19] H. Shao, X. Jiang, P. Gantzel, and M. Goodman, *Chem. Biol.*, 1994, **1**, 231-234.
[20] (a) G. M. Blackburn and J. D. Plackett, *J. Chem. Soc. Perkin Trans. 2*, 1973, 981-985; (b) G. M. Blackburn and J. D. Plackett, *J. Chem. Soc. Perkin Trans. 2*, 1972, 1366-1371.
[21] The Fes protein is composed of the SH2 and kinase domain regions (residues 448 to 822 of human Fes (NP001996)): P. Filippakopoulos, M. Kofler, O. Hantschel, G. D. Gish, F. Grebien, E. Salah, P. Neudecker, L. E. Kay, B. E. Turk, G. Superti-Furga, T. Pawson, and S. Knapp, *Cell*, 2008, **134**, 793-803.
[22] E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. D. Dillon, D. A. Bachovchin, K. Mowen, D. Baker, and B. F. Cravatt, *Nature*, 2010, **468**, 790-795.
[23] T. Böttcher and S. A. Sieber, *Medchemcomm*, 2012, **3**, 408-417.
[24] A. Aitken and M. Learmonth, in *The protein protocols handbook*, ed. J. M. Walker, Humana Press, Totowa, New Jersey, 2nd edn., 2002, vol. IV, pp. 455–456.
[25] J. H. Lee, A. M. Meyer, and H.-S. Lim, *Chem. Commun.*, 2010, **46**, 8615-8617.
[26] (a) L. S. Simpson and T. Kodadek, *Tetrahedron Lett.*, 2012, **53**, 2341-2344; (b) R. Kemperman, A. Kuipers, H. Karsens, A. Nauta, O. Kuipers, and J. Kok, *Appl. Environ. Microbiol.*, 2003, **69**, 1589-1597.
[27] Reactions with cyanogen bromide can lead to side products including oxidation of cysteine to cysteic acid. In addition to the requirement of Met, an often undesirable amino acid in peptide synthesis, side reactions with Ser and Thr residues C-terminal to Met prevent linearization: (a) R. Kaiser and L. Metzka, *Anal. Biochem.*, 1999, **266**, 1-8; (b) B. Smith, in *The Protein Protocols Handbook*, ed. J. M. Walker, Humana Press, Totowa, New Jersey, 1996, pp. 369–373.
[28] D. F. Hunt, J. R. Yates, J. Shabanowitz, S. Winston, and C. R. Hauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 6233-6237.
[29] BioAnalyst v. 1.1 (AB Sciex, Foster City, CA, USA).
[30] J. Qin, C. J. Herring, and X. Zhang, *Rapid Commun. Mass Spectrom.*, 1998, **12**, 209-216.
[31] Whereas linear peptides can be identified by a combination of molecular weight and CID/MS^2, cyclic peptides lack a free N-terminus and cannot be readily sequenced: (a) W.-T. Liu, J. Ng, D. Meluzzi, N. Bandeira, M. Gutierrez, T. L. Simmons, A. W. Schultz, R. G. Linington, B. S. Moore, W. H. Gerwick, P. A. Pevzner, and P. C. Dorrestein, *Anal. Chem.*, 2009, **81**, 4200-4209; (b) M. A. Kelly, T. J. McLellan, and P. J. Rosner, *Anal. Chem.*, 2002, **74**, 1-9.
[32] (a) V. S. Fluxa and J.-L. Reymond, *Bioorg. Med. Chem.*, 2009, **17**, 1018-1025; (b) S. H. Joo, Q. Xiao, Y. Ling, B. Gopishetty, and D. Pei, *J. Am. Chem. Soc.*, 2006, **128**, 13000-13009; (c) S. Li, N. Marthandan, D. Bowerman, H. R. Garner, and T. Kodadek, *Chem. Commun.*, 2005, 581-583; (d) S. M. Williams and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 1039-1054; (e) J. E. Redman, K. M. Wilcoxen, and M. R. Ghadiri, *J. Comb. Chem.*, 2003, **5**, 33-40;
[33] (a) P. Zhao, R. Viner, C. F. Teo, G.-J. Boons, D. Horn, and L. Wells, *J. Proteome Res.*, 2011, **10**, 4088-4104; b) G. L. Juskowiak, C. J. McGee, J. Greaves, and D. L. Van Vranken, *J. Comb. Chem.*, 2008, **10**, 726-731; (c) D. A. Annis, N. Nazef, C.-C. Chuang, M. P. Scott, and H. M. Nash, *J. Am. Chem. Soc.*, 2004, **126**, 15495-15503; (d) R. N. Zuckermann, J. M. Kerr, M. A. Siani, S. C. Banville, and D. V Santi, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 4505-4509.
[34] (a) D. Wu, J. E. Sylvester, L. L. Parker, G. Zhou, and S. J. Kron, *Biopolymers*, 2010, **94**, 475-486; (b) D. G. Udugamasooriya and M. R. Spaller, *Biopolymers*, 2008, **89**, 653-667.
[35] H. Mohimani, Y.-L. Yang, W.-T. Liu, P.-W. Hsieh, P. C. Dorrestein, and P. A. Pevzner, *Proteomics*, 2011, **11**, 3642-3650.

