Organic & Biomolecular Chemistry

This article is part of the

OBC 10th anniversary

themed issue

All articles in this issue will be gathered together online at

www.rsc.org/OBC10

Organic & Biomolecular Chemistry

Dynamic Article Links
View Online

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 5783

www.rsc.org/obc

COMMUNICATION

Selective thioether macrocyclization of peptides having the N-terminal 2-chloroacetyl group and competing two or three cysteine residues in translation†‡

Kazuhiro Iwasaki,[a] Yuki Goto,[b] Takayuki Katoh[b] and Hiroaki Suga*[b]

Received 11th February 2012, Accepted 29th February 2012
DOI: 10.1039/c2ob25306b

The mode of thioether macrocyclization of peptides containing an N-terminal 2-chloroacetyl group and two or three competing cysteine residues at downstream positions has been extensively studied, leading to a strategy for designated formation of overlapping-bicyclic peptides or dumbbell-type bicyclic peptides.

Naturally occurring peptides isolated from various organisms often have macrocyclic structures.[1] These structural features rigidify their tertiary structure, often granting them high affinities to the binding partners or targets. Unlike inter-sidechain disulfide bonds mostly found in ribosomally synthesized proteins, macrocyclization in such peptides takes place in various ways, *i.e.* head-to-tail, head-to-sidechain, sidechain-to-tail, inter-sidechain and their combinations,[2] giving not only their structural diversities but also greater physiological stability.[3] Since the formation of these unique bonds does not occur spontaneously, post-translational modification enzymes or enzymes in a part of non-ribosomal peptide synthetase are involved in the transformation.[4] On the other hand, as such enzymes generally have a preference toward certain sets of peptide sequences and/or particular recognition motifs upstream and/or downstream of the target sequences, their versatility for the synthesis of diverse and artificial macrocyclic peptides could be limited.

To circumvent this limitation, we recently devised a methodology[5] for spontaneous thioether macrocyclization of ribosomally expressed peptides using Flexible *In-vitro* Translation (FIT) system.[6] In this system, an artificial amino acid bearing a 2-chloroacetyl (ClAc) group on the α-amino group was incorporated by initiation codon reprogramming at the N-terminus, and reacts with sulfhydryl group in a cysteine (C) residue at a downstream position (Fig. 1a). Although this head-to-sidechain macrocyclization chemistry itself has been known for the chemical synthesis of peptides,[7] the above method using the FIT system represents the first example of thioether macrocyclization of ribosomally expressed peptides. Most importantly, this methodology allows us to readily construct various sequences and lengths of thioether macrocyclic peptides by simply designing their mRNA sequences which are transcribed from the corresponding synthetic DNAs *in vitro*.[5] Moreover, we have recently constructed such genetically encoded libraries of thioether macrocyclic peptides with a complexity of over 10^{12} and applied them to the discovery of inhibitors against various enzyme targets, *e.g.* E6AP,[8] Akt2,[9] SIRT2,[10] by integration with the mRNA display technique,[11] referred to as RaPID (Random non-standard Peptide Integrated Discovery) system as a whole.

A typical design of the above libraries is to have an *N*-ClAc-amino acid encoded by AUG as an initiator, random elongator amino acid sequences encoded by $(NNK)_n$ (K = U or G, n = 7–15), $(NNU)_n$, or $(NNC)_n$, followed by a cysteine (C) encoded

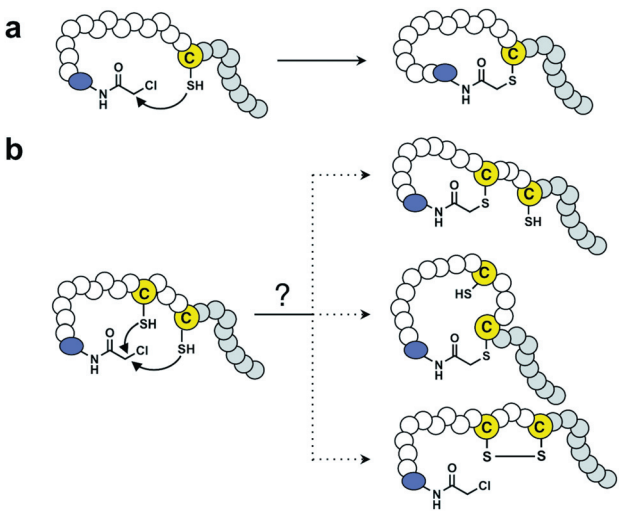

Fig. 1 Schematic representation of peptide macrocyclization closed by a thioether bond. (**a**) Peptide with N-terminal ClAc group that reacts with a cysteine (C) residue. (**b**) Three possible cyclization pathways in *N*-ClAc-peptides with two cysteine residues.

[a]*Department of Chemistry and Biotechnology, Graduate School of Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[b]*Department of Chemistry, Graduate School of Science, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan. E-mail: hsuga@chem.s.u-tokyo.ac.jp*

†This article is part of the *Organic & Biomolecular Chemistry* 10th Anniversary issue.

‡Electronic supplementary information (ESI) available: Materials, Experimental procedures, MS data, and MS/MS data. See DOI: 10.1039/c2ob25306b

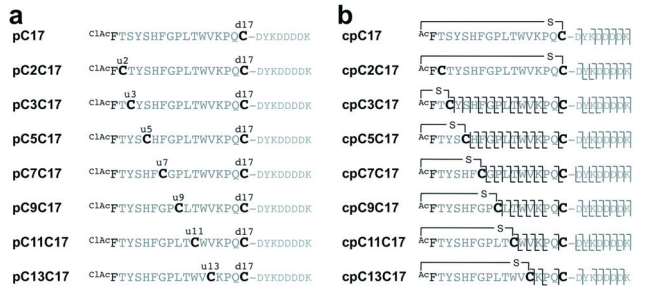

Fig. 2 Peptides and their fragmentation profiles observed in MALDI-TOF/TOF mass spectrometry. (**a**) Sequences of linear peptides expressed in this study. (**b**) Observed fragment peaks in each cyclized peptide and putative thioether bond formation based on the fragmentation profile. L shape and rotated L shape denote observations of y-ion and b-ion peaks, respectively.

by UGC. As a result, the major fraction of expressed peptides would cyclize between the N-terminus and the designated C residue at the C-terminus. However, during the course of such selection studies, we encountered active peptide sequences containing two (or possibly more) cysteine residues, one of which appeared in the random region encoded by UGU or UGC. In such peptides (Fig. 1b), both C residues could potentially react with the N-terminal ClAc group to form a thioether bond,[8] but it was unclear which bond would be preferentially or even selectively formed in such peptide sequences (Fig. 1b, top *vs.* middle). Alternatively, a disulfide bond between two C residues could possibly compete with the thioether bond formation (Fig. 1b, bottom). To understand a general rule of macrocyclization in such peptides, here we have designed experiments to express model peptides varying the position of an 'upstream Cys' (^{u}C) residue in addition to the designated 'downstream Cys' (^{d}C) residue.

We used a FIT system customized for the initiation codon reprogramming, where an initiator tRNA charged with *N*-ClAc-phenylalanine (^{ClAc}Phe) by means of flexizyme was supplemented to a Met-deficient cell-free translation system. We designed seven mRNA sequences encoding 25-mer model peptides (Fig. 2a), each of which was initiated with ^{ClAc}Phe, elongated with 15 residues containing ^{u}C and ^{d}C, and further elongated with a Flag tag (DYKDDDDK, where D, Y, K are aspartate, tyrosine, and lysine; it is used for not only an affinity purification tag but also for facilitating ionization of expressed peptides in mass spectrometry). In these peptides, ^{u}C was located in the 2nd position or an odd position from the 3rd to the 13th position along with a fixed ^{d}C at the 17th position (pC2C17–pC13C17, where the former and latter number denotes the position of ^{u}C and ^{d}C residue and p strands for peptide). Thus, each peptide would possibly produce three macrocyclic structures with a thioether bond between the N-terminus and ^{u}C or ^{d}C, or an ^{u}C–^{d}C disulfide bond. As a control, an mRNA sequence encoding a simple thioether macrocyclic peptide was also prepared (pC17).

All mRNA sequences expressed the encoded peptide with a loss of approximately 36 Da from the molecular mass of the full-length peptide, *i.e.* a loss of HCl, as a major product peak confirmed by MALDI-TOF (Fig. S1‡). This result indicated that the thioether bond formation preferentially occurred over the disulfide bond formation. The respective peptides were then subjected to the fragmentation by MALDI-TOF/TOF, of which the individual peaks were assigned to the expected fragments. Interestingly, two peptides, cpC17 and cpC2C17 (c stands for cyclic) were poorly fragmented at the region between F1 and ^{d}C17 (Fig. S2‡). On the other hand, cpC3C17–cpC13C17 afforded various peaks of fragments, supposedly originating from the linear region when macrocyclization took place between the N-terminus and ^{u}C; whereas, fragment of the thioether macrocyclic region was not found similar to that observed for cpC17 (Fig. S3‡). These results consistently suggested that macrocyclization of pC3C17–pC13C17 took place between the N-terminus and ^{u}C, giving the respective structures consisting of a macrocyclic head and a linear tail with a free sulfhydryl on ^{d}C17. In contrast, macrocyclization of pC2C17 predominantly took place between the N-terminus and ^{d}C17, giving the macrocyclic structure with a free sulfhydryl on ^{u}C2 (Fig. 2b). This selective macrocyclization could be attributed to the fact that the nucleophilic attack of the sulfhydryl group of ^{u}C2 on the α-carbon of the *N*-ClAc group required the formation of a constrained 9-membered ring, which might be unfavourable over macrocyclization between the N-terminus and ^{d}C17.

To solidify the above hypothesis, we expressed cpC3C17 peptide containing a single ester bond (cpC3e12C17; e denoting the ester bond) where tryptophan at the 12th position (W12) was replaced with phenyllactic acid (^{HO}F) by reprogramming of its codon (Fig. S4a‡).[12] As a control, we also expressed cpC17 with the same ester substitution at the 12th position (Fig. S4b‡). MALDI-TOF analysis of each peptide confirmed the production of macrocyclic peptide-ester hybrid as a single major product (Fig. S4c and d‡). Upon subjecting to alkaline hydrolysis, the ester bond was selectively cleaved to yield a product or products as follows: in cpC3e12C17, two fragment peaks were observed, each of which was assigned to the thioether macrocyclic peptide with the carboxylate tail (**1**) and the linear tail peptide (**2**) cleaved at ^{HO}F, respectively (Fig. S4c‡); in a control peptide cpe12C17, on the other hand, a single peak with the molecular weight corresponding to the hydrolysis product at the 12th position (**3**) was detected as a single piece of sequence (Fig. S4d‡). These results were consistent with our hypothesis that the macrocyclization of pC3C17–pC13C17 took place between the N-terminus and ^{u}C.

With all knowledge together, we designed bicyclic peptides consisting of a head-to-sidechain thioether bond and an inter-sidechain disulfide bond. We first demonstrated expression of a peptide, referred to as pC2C10C17, where leucine at the 10th position (L10) in pC2C17 was replaced with a 'middle Cys' (^{m}C10) residue. Because ^{u}C2 should not react with the N-terminal ClAc group, we expected that the sulfhydryl group of ^{m}C10 would spontaneously react with the N-terminus, leaving two thiol-free cysteine residues, ^{u}C2 and ^{d}C17; then these two residues would form a disulfide bond under mild oxidation conditions, yielding an overlapping-bicyclic backbone structure (Fig. 3a). In fact, the mass spectrum of the expressed pC2C10C17 after air-oxidization showed a peak corresponding to the expected bicyclic structure (Fig. 3b, pC2C10C17-oxi). Furthermore, the tandem mass analysis of cpC2C10C17-oxi by MALDI-TOF/TOF showed fragment peaks only corresponding

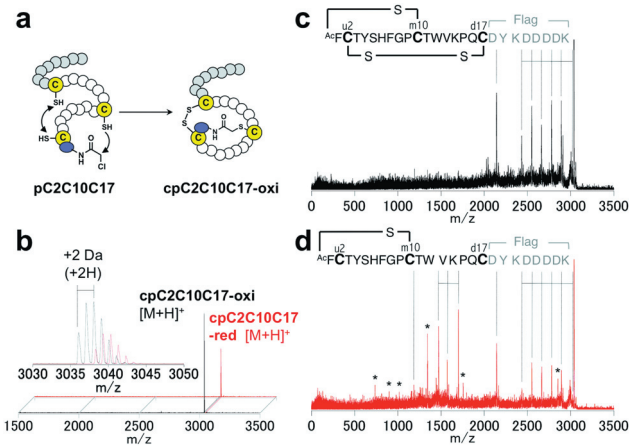

Fig. 3 Expression of an overlapping-bicyclic peptide consisting of a thioether bond and a disulfide bond. (a) Schematic representation of macrobicyclic peptides based on the pC2C10C17 scaffold. Because C2 is less susceptible to the reaction with the *N*-ClAc group, C10 predominantly forms a thioether bond and the remaining C2 and C17 form a disulfide bond. (b) MALDI-TOF spectra of cpC2C10C17 with two forms. Mild air-oxidation of the expressed peptide yielded the peptide, cpC2C10C17-oxi, that was consistent with the molecular mass containing a thioether bond and a disulfide bond (black spectrum). Its TCEP treatment yielded a mass with an additional 2 Da (cpC2C10C17-red), suggesting that the disulfide bond was reduced (red spectrum). The inset figure shows an expanded 3030–3050 *m/z* area of the superimposed spectra of cpC2C10C17-oxi (Cal. 3036.23, Obs. 3036.08) and cpC2C10C17-red (Cal. 3038.24, Obs. 3038.32). (c) MALDI-TOF/TOF spectrum of cpC2C10C17-oxi. (d) MALDI-TOF/TOF spectrum of cpC2C10C17-red. Assigned b-ion peaks are shown in the figure. Peaks labeled with asterisks are assigned as y-ion peaks.

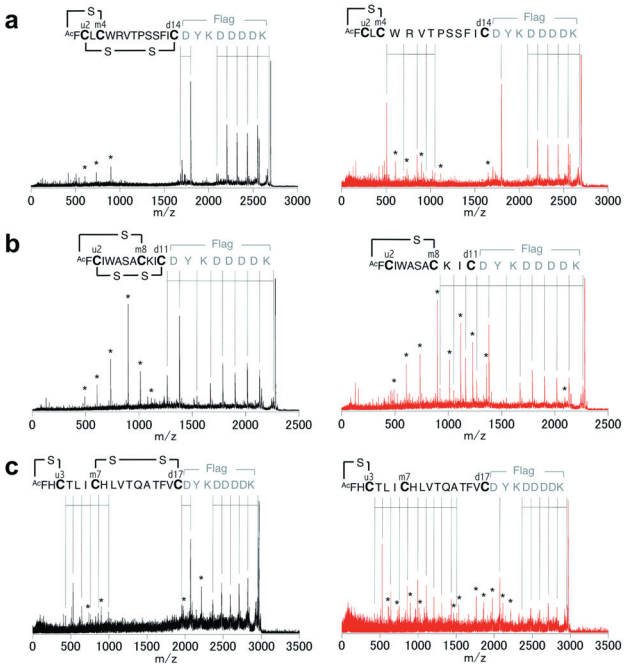

Fig. 4 MALDI-TOF/TOF spectra of overlapping- or dumbbell-type macrocyclic peptides. (a) cpC2C4C14, (b) cpC2C8C11, and (c) cpC3C7C17. Mild air-oxidation of the respective expressed peptide yielded the bicyclic structure that was consistent with the molecular mass containing a thioether bond and a disulfide bond (spectra of peptide-oxi shown in black). Upon treatment with TCEP, the molecular mass of each peptide-oxi increased by 2 Da, suggesting that the disulfide bond was reduced to yield monocyclic peptide (spectra of peptide-red shown in red). Assigned b-ion peaks are shown in the figure. Peaks labeled with asterisks are assigned as y-ion peaks.

to the linear Flag region (Fig. 3c), indicating that the desired overlapping-bicyclic structure was formed. Upon reduction treatment of pC2C10C17-oxi with tris(2-carboxyethyl)phosphine (TCEP), the molecular mass of pC2C10C17-oxi increased by approximately 2 Da, implying that the disulfide bond was reduced to form pC2C10C17-red (Fig. 3b, pC2C10C17-red). Tandem mass analysis of pC2C10C17-red yielded additional peaks corresponding to some fragmentations in the region between ^{m}C10 and ^{d}C17 (Fig. 3d). This result gave solid evidence that the designed overlapping-bicyclic peptide consisting of the head-to-sidechain and inter-sidechain bonds was produced.

To demonstrate the designer generality of this approach, we also expressed ^{u}C2-peptides varying in the position of ^{m}CX (X = 4, 5, 6, or 8) in the combination of ^{d}CY (Y = 7, 10, 11, or 14) with arbitrary sequence compositions and lengths. In all cases, production of the desired bicyclic peptide was confirmed by MALDI-TOF/TOF (Fig. 4a for X = 4 and Y = 14, and b for X = 8 and Y = 11; see also Fig. S5a, b, c, and d‡ for other combinations). Moreover, ^{u}C3-peptides having two different positions of ^{m}CX (X = 7, and 13) were also expressed, yielding dumbbell-type peptides (Fig. 4c for X = 7; see also Fig. S5e‡ for X = 13). Taken together, this designer bicyclization strategy enables us to construct various bicyclic peptides with desired sequence compositions.

Conclusions

Here we have shown selective macrocyclization of peptides having the N-terminal ClAc group with two competing C residues expressed in translation machinery under the reprogrammed genetic code. The nearest ^{u}C (upstream C) residue generally dictates the spontaneous thioether bond formation *via* thiol-nucleophilic attack on the α-carbon of the ClAc group, thus forming a macrocyclic structure and leaving a ^{d}C (downstream C) residue unreacted. However, there is one exception where ^{u}C embedded at position 2 (^{u}C2) does not react with the ClAc group and instead ^{d}C residue predominantly reacts with the ClAc group, leaving the ^{u}C residue thiol-free. Based on this knowledge, we have devised a strategy for the synthesis of bicyclic peptides in which an additional middle C (^{m}C) residue between the ^{u}C2 and ^{d}C residues is embedded in the sequence, cyclizing the peptide by a thioether bond between ClAc and ^{m}C followed by a disulfide bond formation of ^{u}C2 and ^{d}C. This bicyclization is selective and fairly general independent of sequence compositions. Thus, this strategy is applicable for constructing a library consisting of such constrained bicyclic peptide scaffolds. Particularly, the integration of this library with RaPID system will allow us to rapidly select bicyclic peptides against a chosen protein target. This opens a new avenue to search a unique peptide 3-dimensional sequence space for bioactive peptides.

Acknowledgements

We thank H. Murakami for the discussion and development of the FIT system. This work was supported by a Grants-in-Aid of Japan Society for Promotion of Science (JSPS), the Specially Promoted Research (21000005) and the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization to H.S.; Grants-in-Aid for JSPS Fellows (22–9695) to K.I.; Grants-in-Aid for Young Scientists (B) (22750145) from JSPS and PRESTO, Japan Science and Technology Agency (10206) to Y.G.; and Grants-in-Aid for Young Scientists (B) (22710210) from JSPS to T.K.

Notes and references

1 D. P. Fairlie, G. Abbenante and D. R. March, *Curr. Med. Chem.*, 1995, **2**, 654–686.

2 (*a*) A. Paladini and L. C. Craig, *J. Am. Chem. Soc.*, 1954, **76**, 688–692; (*b*) M. Dreyfuss, E. Harri, H. Hofmann, H. Kobel, W. Pache and H. Tscherter, *Eur. J. Appl. Microbiol.*, 1976, **3**, 125–133; (*c*) A. C. Whyte, B. K. Joshi, J. B. Gloer, D. T. Wicklow and P. F. Dowd, *J. Nat. Prod.*, 2000, **63**, 1006–1009; (*d*) T. Tanaka, E. Tsukuda, M. Nozawa, H. Nonaka, T. Ohno, H. Kase, K. Yamada and Y. Matsuda, *Mol. Pharmacol.*, 1994, **45**, 724–730; (*e*) B. A. Johnson, H. Anker and F. L. Meleney, *Science (Washington, D. C.)*, 1945, **102**, 376–377.

3 D. R. March, G. Abbenante, D. A. Bergman, R. I. Brinkworth, W. Wickramasinghe, J. Begun, J. L. Martin and D. P. Fairlie, *J. Am. Chem. Soc.*, 1996, **118**, 3375–3379.

4 (*a*) D. Schwarzer, R. Finking and M. A. Marahiel, *Nat. Prod. Rep.*, 2003, **20**, 275–287; (*b*) R. Finking and M. A. Marahiel, *Annu. Rev. Microbiol.*, 2004, **58**, 453–488; (*c*) M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.

5 Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 120–129.

6 (*a*) H. Murakami, A. Ohta, H. Ashigai and H. Suga, *Nat. Methods*, 2006, **3**, 357–359; (*b*) Y. Goto, T. Katoh and H. Suga, *Nat. Protocols*, 2011, **6**, 779–790.

7 (*a*) L. Yu, Y. H. Lai, J. V. Wade and S. M. Coutts, *Tetrahedron Lett.*, 1998, **39**, 6633–6636; (*b*) K. D. Roberts, J. N. Lambert, N. J. Ede and A. M. Bray, *Tetrahedron Lett.*, 1998, **39**, 8357–8360; (*c*) F. D. T. Lung, C. R. King and P. P. Roller, *Lett. Pept. Sci.*, 1999, **6**, 45–49; (*d*) S. C. Pero, L. Oligino, R. J. Daly, A. L. Soden, C. Liu, P. P. Roller, P. Li and D. N. Krag, *J. Biol. Chem.*, 2002, **277**, 11918–11926; (*e*) P. Li, M. L. Peach, M. C. Zhang, H. P. Liu, D. J. Yang, M. Nicklaus and P. P. Roller, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 895–899; (*f*) F. M. Brunel and P. E. Dawson, *Chem. Commun.*, 2005, 2552–2554; (*g*) R. Tugyi, G. Mezo, E. Fellinger, D. Andreu and F. Hudecz, *J. Pept. Sci.*, 2005, **11**, 642–649; (*h*) K. D. Roberts, J. N. Lambert, J. Ede and A. M. Bray, *J. Pept. Sci.*, 2006, **12**, 525–532; (*i*) C. J. Porter and J. A. Wilce, *Biopolymers*, 2007, **88**, 174–181; (*j*) K. D. Roberts and N. J. Ede, *J. Pept. Sci.*, 2007, **13**, 811–821; (*k*) S. Jiang, C. Z. Liao, L. Bindu, B. L. Yin, K. W. Worthy, R. J. Fisher, T. R. Burke, M. C. Nicklaus and P. P. Roller, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 2693–2698.

8 Y. Yamagishi, I. Shoji, S. Miyagawa, T. Kawakami, T. Katoh, Y. Goto and H. Suga, *Chem. Biol.*, 2011, **18**, 1562–1570.

9 Y. Hayashi, J. Morimoto and H. Suga, *ACS Chem. Biol.*, 2012, DOI: 10.1021/cb200388k.

10 J. Morimoto, Y. Hayashi and H. Suga, *Angew. Chem., Int. Ed. Engl.*, 2012, DOI: 10.1002/anie.201108118.

11 (*a*) R. W. Roberts and J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 12297–12302; (*b*) N. Nemoto, E. MiyamotoSato, Y. Husimi and H. Yanagawa, *FEBS Lett.*, 1997, **414**, 405–408.

12 (*a*) In addition to genetic code reprogramming, a hydroxy acid has been incorporated into amber codon in the following papers: A. Ohta, H. Murakami, E. Higashimura and H. Suga, *Chem. Biol.*, 2007, **14**, 1315–1322; (*b*) J. A. Ellman, D. Mendel and P. G. Schultz, *Science (Washington, D. C.)*, 1992, **255**, 197–200; (*c*) P. M. England, H. A. Lester and D. A. Dougherty, *Biochemistry*, 1999, **38**, 14409–14415; (*d*) S. W. Millward, T. T. Takahashi and R. W. Roberts, *J. Am. Chem. Soc.*, 2005, **127**, 14142–14143.

