ChemComm

COMMUNICATION

Cite this: *Chem. Commun.*, 2017, 53, 9846

Received 18th July 2017,
Accepted 11th August 2017

DOI: 10.1039/c7cc05548j

rsc.li/chemcomm

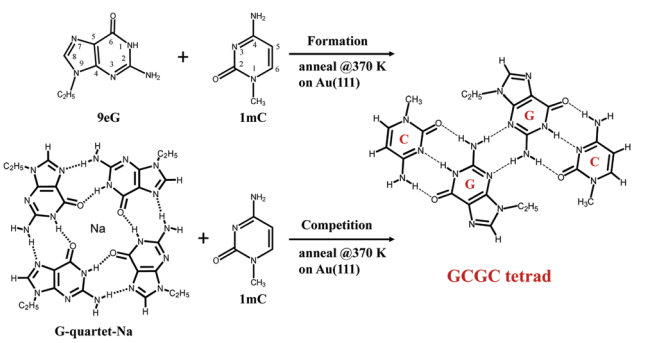
Real-space evidence of the formation of the GCGC tetrad and its competition with the G-quartet on the Au(111) surface†

Yuanqi Ding, Lei Xie, Chi Zhang and Wei Xu *

From the interplay of high-resolution scanning tunneling microscopy (STM) imaging and density functional theory (DFT) calculations, we show the first real-space evidence of the formation of GCGC tetrad on an Au(111) surface, and further investigate its competition with the well-known G-quartet with the aid of NaCl under ultrahigh vacuum (UHV) conditions.

DNA quadruplexes, as a kind of non-canonical DNA structure formed by guanine (G)-rich nucleic acid sequences, play a significant role in several biological functions and processes.[1] This four-stranded architecture is mainly formed by stacked G-quartets stabilized by cations,[2] which has been well studied.[3] Meanwhile, many studies show that some unusual tetrads are also involved and located at the end of DNA quadruplexes, such as stable mixed tetrads ATAT (adenine and thymine tetrad) and GCGC (guanine and cytosine tetrad) stacked above G-quartets.[4–8] Among others, GCGC tetrads have been found in quadruplex structures *in vitro* formed by sequences containing GGGC tandem repeats of adeno-associated viral gene[6] and human chromosome 19,[4,9] and also the CGG triplet repeats of fragile X syndrome disease gene.[7,10,11] Owing to its biological relevance as well as the subtle relationship with the well-known G-quartet in DNA quadruplexes, the GCGC tetrad has thus attracted tremendous interest. Over the past few decades, experimental studies of the GCGC tetrad have been confined in solution systems,[4–8,10,12–15] and the supposed tetrad structures of GCGC were mostly determined by NMR spectroscopy,[4–8,12–15] X-ray diffraction[16,17] and theoretical calculations.[18–21] Although the G-quartet and Watson–Crick GC base pair structures have been successfully introduced to the surface science community, and extensively studied by scanning tunneling microscopy (STM),[22–35] to our knowledge, the real-space evidence of GCGC tetrads has not been demonstrated previously. Thus, it is of utmost interest to construct a system under well-controlled ultrahigh vacuum (UHV) conditions with the aim of forming GCGC tetrads on a surface and gain the atomic-scale structure by STM imaging. Moreover, such a study may provide us a system to unravel some biologically relevant issues, *e.g.*, the competition between the GCGC tetrad and the G-quartet.

In this study, we choose modified DNA bases of 9-ethylguanine (9eG) and 1-methylcytosine (1mC) molecules together with the Au(111) surface as a model system. Both base derivatives are modified at the same sites as that of the natural nucleosides (*cf.* Scheme 1) and have been proved to form a 1 : 1 complex in the crystal structure.[17] The Au(111) surface is employed as a template to ensure that the molecules adopt a flat adsorption geometry to facilitate the potential formation of a planar GCGC tetrad (9-ethylguanine and 1-methylcytosine tetrad). Herein, from the interplay of high-resolution STM imaging and density functional theory (DFT) calculations, we show that (i) the formation of chain structures composed of the GCGC tetrad as the elementary motif is achieved by the deposition of 9eG and 1mC molecules on Au(111) and annealing at 370 K as depicted in Scheme 1; (ii) further annealing of the chain structures with NaCl results in the formation of network structures where the



Scheme 1 Schematic illustration of the formation of a GCGC tetrad (9-ethylguanine and 1-methylcytosine tetrad) structure and the competition with the G-quartet-Na (9-ethylguanine-quartet-Na complex) structure on the Au(111) surface.

Interdisciplinary Materials Research Center, Tongji-Aarhus Joint Research Center for Nanostructures and Functional Nanomaterials, College of Materials Science and Engineering, Tongji University, Shanghai 201804, P. R. China.
E-mail: xuwei@tongji.edu.cn

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cc05548j

individual GCGC tetrads are usually linked together by both Na and Cl without perturbing the intra-tetrad hydrogen bonds, which demonstrates the robustness of such a GCGC tetrad structure; (iii) the deposition of 1mC molecules on the G-quartet-Na (9-ethylguanine-quartet-Na complex) precovered surface and then annealing at 370 K, interestingly, result in the transformation to the GCGC tetrad networks (*cf.* Scheme 1), which suggests that the GCGC tetrad is thermodynamically more stable than the G-quartet-Na when 1mC molecules are provided; and (iv) a comparative study shows that the direct annealing of the mixture of 9eG, 1mC and NaCl also results in the formation of the GCGC tetrad networks, which further illustrates the competition between the G-quartet-Na and the GCGC tetrad motifs on the surface. These findings demonstrate the first real-space experimental evidence of the formation of the GCGC tetrad on the Au(111) surface, and prove its stability in comparison with the G-quartet-Na structure.

The deposition of 9eG and 1mC molecules (at a 9eG/1mC ratio of ~1:1, the STM images of the consequence at other stoichiometric ratios are shown in Fig. S1a and b, ESI†) on the Au(111) surface held at room temperature (RT) and further annealing at 370 K result in the formation of a chain structure without certain registry with the substrate as shown in Fig. 1a and b (at low and high coverage, respectively). This chain structure is obviously different from the self-assembled structures of 1mC molecules (Fig. S2a, ESI†) and 9eG molecules (Fig. S2b, ESI†), which should be a mixed structure of the two molecules. From the buckling sites and defects shown in Fig. 1a, we identify that such chain structures are usually composed of one kind of tetrad motif. A closer inspection of the chains (*cf.* Fig. 1c) allows us to distinguish that some tetrads (indicated by white contours) can be separated by the 9eG dimers, which further confirms that such a tetrad should be the elementary structural motif. From the high-resolution STM image of the single chain (Fig. 1d), we identify that such a tetrad is composed of two 9eG molecules and two 1mC molecules. Each individual 9eG molecule is resolved as a triangle part (*i.e.* the guanine moiety) with a bright round protrusion (*i.e.* tilted ethyl group),[28] while each 1mC molecule is shown as an ellipse (as indicated by G and C notations, respectively, in Fig. 1d). Based on the above assignment, we have performed DFT calculations to gain the atomic-scale model of such a GCGC tetrad structure. From the DFT-optimized model superimposed on the corresponding high-resolution STM image (Fig. 1e), we distinguish that the tetrad is formed by two conventional Watson–Crick GC pairs linked together by additional double NH···N hydrogen bonds between 9eG molecules (as depicted by blue dashed lines in Fig. 1e). The corresponding simulated STM image (Fig. 1f) shows a good agreement with the experimental one. The GCGC tetrads are further linked with the neighboring ones *via* double NH···O hydrogen bonds forming chain structures as shown in Fig. S3 (ESI†). Further annealing such a sample to 420 K (the desorption temperature of 1mC molecules) leads to the formation of the 9eG island structure (see Fig. S4, ESI†).[28,33]

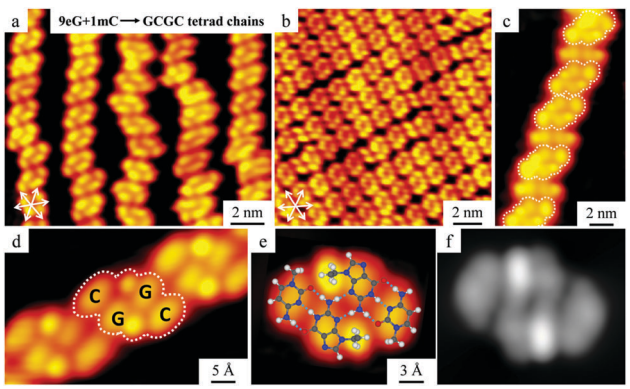


Fig. 1 Formation of GCGC tetrad chains after deposition of 9eG and 1mC molecules on Au(111) and annealing at 370 K. Large-scale STM images at a low (a) and high (b) coverage showing the GCGC tetrad chains. (c) STM image of a single chain with GCGC tetrads separated by 9eG dimers. (d) Close-up STM image of a single chain allowing us to identify the individual 9eG and 1mC molecules (as indicated by G and C notations). (e) High-resolution STM image of the GCGC tetrad superimposed with the DFT-optimized model on Au(111) (the surface is removed for clear presentation). Hydrogen bonds are depicted by blue dashed lines. (f) The simulated STM image of the GCGC tetrad at a bias voltage of 1.2 V. The GCGC motifs are depicted by white contours in (c) and (d). H: white; C: gray; N: blue; O: red.

As previously shown in the literature,[36–39] also specifically our own work,[34,35] 9eG molecules cannot form the empty quartet structure without cations, and NaCl can be employed as a reactant to effectively recognize and interact with the 9eG molecules resulting in the formation of the G-quartet-Na structure, it is then naturally interesting to introduce NaCl to this GCGC system to detect the stability of GCGC tetrads. After subsequent deposition of NaCl on the GCGC tetrad chain precovered surface (Fig. 2a) and further annealing at 370 K, surprisingly, the GCGC tetrad chains transform into a network structure as shown in Fig. 2b. More importantly, from the close-up STM image (Fig. 2c), we identify that such a network structure is composed of the same GCGC tetrad as the elementary motif (as depicted by the white contours), and an isolated GCGC tetrad is also observed as highlighted by the green contour. According to the previous findings,[40–42] halogen atoms usually appear as bright protrusions in STM images, we thus temporarily assign the bright spots located between the GCGC tetrads in the network to the Cl. Note that Na is normally invisible in most tip states, which is also the case for the G-quartet-Na structure shown before.[34,35] To confirm the necessity of Na in the formation of such a network structure, a comparative experiment has been performed by vaporing iodine (as a substitute to Cl due to its facile deposition method and the similar properties) to the GCGC tetrad chain precovered surface. As shown in Fig. S5 (ESI†), we still observe the formation of chain structures composed of the same GCGC tetrads organized by iodine, which is characteristically different from what we have shown in Fig. 2b. The different arrangement suggests the essential role of Na in the formation of the GCGC tetrad networks. As experienced with our previous studies of DNA bases and alkali metals,[31,34] we speculate that Na interacts with 9eG molecules at N7 and O6 sites (as marked in Scheme 1). Based on that, we performed DFT calculations, and the optimized gas-phase model of the GCGC tetrad networks is superimposed on the corresponding STM image as shown in Fig. 2d. From the model, we identify that the GCGC tetrads are linked together by both Na and Cl without perturbing the intra-tetrad

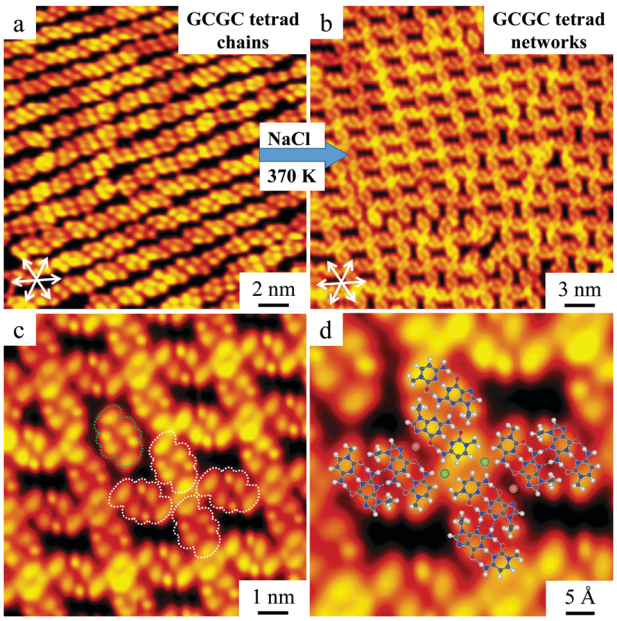


Fig. 2 STM images showing the transformation from (a) GCGC tetrad chains to (b) GCGC tetrad networks on Au(111) after interaction with NaCl. (c) The close-up STM image of the network structure with individual GCGC tetrads depicted by white contours. An isolated GCGC tetrad is highlighted by the green contour. (d) High-resolution STM image superimposed with the DFT-optimized gas-phase model. Hydrogen bonds are depicted by blue dashed lines. H: white; C: gray; N: blue; O: red; Na: pink; Cl: green.

hydrogen bonds, which demonstrates the robustness of such a GCGC tetrad structure.

Since the G-quartet and the GCGC tetrad are both fundamental building blocks of DNA quadruplexes,[3] and the GCGC tetrad has been demonstrated to have great stability as shown above, it would be generally interesting to further explore the competition between the G-quartet and the GCGC tetrad on the surface. To do so, we have performed the following comparative experiments. We first fabricate the G-quartet-Na network according to our previous procedure[34] by the deposition of 9eG molecules and NaCl on the surface and subsequent annealing to 350 K as shown in Fig. 3a and b. And then, we deposit 1mC molecules on the G-quartet-Na covered surface followed by annealing to 370 K. Surprisingly, we observe the structural transformation from the G-quartet-Na network to the GCGC tetrad network as shown in Fig. 3c. Further annealing the sample to the desorption temperature of 1mC molecules at 420 K results in the formation of isolated G-quartet-Na as depicted by blue contours in Fig. 3d, which also proves the existence of Na in the GCGC tetrad network structure. Based on the above results, the reversible transformation between the G-quartet-Na and the GCGC tetrad is achieved on the surface. More importantly, it suggests that the formation of the GCGC tetrad is thermodynamically more favorable than that of the G-quartet-Na with the existence of 1mC molecules on the Au(111) surface.

Finally, simultaneous deposition of 9eG and 1mC molecules together with NaCl on the surface and followed by annealing to 370 K also result in the formation of the GCGC tetrad network as shown in Fig. 4a and b as expected. Such a control experiment

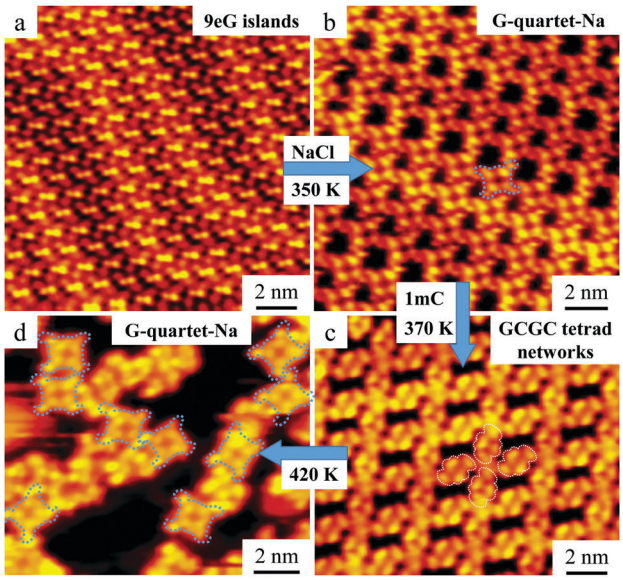


Fig. 3 STM images showing the transformations from the 9eG island *via* the G-quartet-Na network to the GCGC tetrad network, and finally back to the G-quartet-Na network. (a and b) STM images showing the formation of the G-quartet-Na network after deposition of NaCl on the 9eG island covered surface and annealing to 350 K on Au(111). (c) STM image showing the formation of the GCGC tetrad network after deposition of 1mC molecules on the G-quartet-Na network covered surface and annealing to 370 K. (d) STM image showing the transformation back to the G-quartet-Na structure after further annealing the GCGC tetrad network to 420 K. The G-quartet-Na network and the GCGC tetrad are depicted by blue and white contours, respectively.

further confirms the competition scenario between the G-quartet-Na and the GCGC tetrad on the Au(111) surface.

To quantify the competition between the G-quartet-Na and the GCGC tetrad, we performed calculations on the G-quartet-Na network and the GCGC tetrad network in the gas phase, and calculated the binding energy (E_b) to characterize their relative stabilities. E_b is defined as the total energy of the whole relaxed system (including the hydrogen bonds and electrostatic interactions in the formed molecular structures) minus the total energies of all of the individual relaxed components. That is, $E_b = E_{sys} - E_{com}$, where E_{sys} represents the total energy of the whole system, and E_{com} represents the total energy of the individual relaxed components. The E_b value of the G-quartet-Na network is calculated to

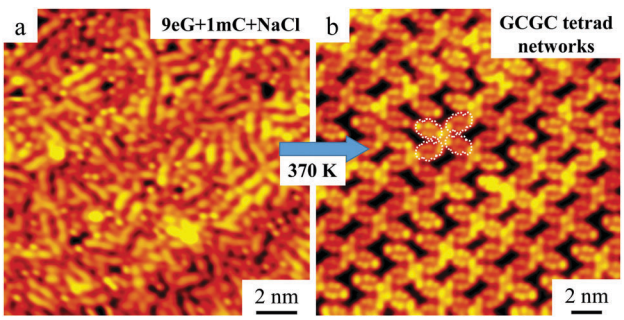


Fig. 4 STM images showing the formation of the GCGC tetrad networks after annealing the mixture of 9eG, 1mC and NaCl to 370 K on Au(111).

be 1.51 eV per molecule, and the E_b value of the GCGC tetrad network (including the interactions of Na and Cl with molecules) is 2.76 eV per molecule. From the above results, we draw the conclusion that the GCGC tetrad network is thermodynamically more stable than the G-quartet-Na network when 1mC molecules are provided.

In conclusion, by the combination of sub-molecularly resolved STM imaging and DFT calculations, we have presented the first real-space experimental evidence of the formation of GCGC tetrads on Au(111) and further revealed its competition with the G-quartet-Na structure. Such a surface science study on the two fundamental building blocks of DNA quadruplexes may help to increase our fundamental understandings of DNA quadruplex structures.

The authors acknowledge financial support from the National Natural Science Foundation of China (21473123 and 21622307).

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 H. Arthanari and P. H. Bolton, *Chem. Biol.*, 2001, **8**, 221.
2 M. Gellert, M. N. Lipsett and D. R. Davies, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013.
3 J. T. Davis, *Angew. Chem., Int. Ed.*, 2004, **43**, 668.
4 S. Bouaziz, A. Kettani and D. J. Patel, *J. Mol. Biol.*, 1998, **282**, 637.
5 N. Escaja, E. Pedroso, M. Rico and C. González, *J. Am. Chem. Soc.*, 2000, **122**, 12732.
6 A. Kettani, S. Bouaziz, A. Gorin, H. Zhao, R. A. Jones and D. J. Patel, *J. Mol. Biol.*, 1998, **282**, 619.
7 A. Kettani, R. A. Kumar and D. J. Patel, *J. Mol. Biol.*, 1995, **254**, 638.
8 N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin and D. J. Patel, *J. Mol. Biol.*, 2001, **312**, 1073.
9 R. M. Kotin, R. M. Linden and K. I. Berns, *EMBO J.*, 1992, **11**, 5071.
10 J. M. Darlow and D. R. F. Leach, *J. Mol. Biol.*, 1998, **275**, 3.
11 H. Moine and J. L. Mandel, *Science*, 2001, **294**, 2487.
12 N. Escaja, I. Gómez-Pinto, E. Pedroso and C. González, *J. Am. Chem. Soc.*, 2007, **129**, 2004.
13 M. Frańska, *Int. J. Mass Spectrom.*, 2015, **387**, 83.
14 S. Metzger and B. Lippert, *J. Am. Chem. Soc.*, 1996, **118**, 12467.
15 N. G. Williams, L. D. Williams and B. R. Shaw, *J. Am. Chem. Soc.*, 1989, **111**, 7205.
16 G. A. Leonard, S. Zhang, M. R. Peterson, S. J. Harrop, J. R. Helliwell, W. B. Cruse, B. L. d'Estaintot, O. Kennard, T. Brown and W. N. Hunter, *Structure*, 1995, **3**, 335.
17 E. J. O'Brien, *Acta Crystallogr.*, 1967, **23**, 92.
18 J. Gu, J. Wang and J. Leszczynski, *J. Am. Chem. Soc.*, 2004, **126**, 12651.
19 J. Gu and J. Leszczynski, *J. Phys. Chem. A*, 2000, **104**, 7353.
20 M. Meyer, C. Schneider, M. Brandl and J. Sühnel, *J. Phys. Chem. A*, 2001, **105**, 11560.
21 M. Meyer and J. Sühnel, *J. Biomol. Struct. Dyn.*, 2003, **20**, 507.
22 N. Bilbao, I. Destoop, S. De Feyter and D. González-Rodríguez, *Angew. Chem., Int. Ed.*, 2016, **55**, 659.
23 A. Ciesielski, S. Lena, S. Masiero, G. P. Spada and P. Samorì, *Angew. Chem., Int. Ed.*, 2010, **49**, 1963.
24 A. Ciesielski, R. Perone, S. Pieraccini, G. P. Spada and P. Samorì, *Chem. Commun.*, 2010, **46**, 4493.
25 D. González-Rodríguez, P. G. Janssen, R. Martín-Rapún, I. De Cat, S. De Feyter, A. P. Schenning and E. W. Meijer, *J. Am. Chem. Soc.*, 2010, **132**, 4710.
26 R. Otero, M. Schöck, L. M. Molina, E. Lægsgaard, I. Stensgaard, B. Hammer and F. Besenbacher, *Angew. Chem., Int. Ed.*, 2005, **44**, 2270.
27 R. Otero, W. Xu, M. Lukas, R. E. Kelly, E. Lægsgaard, I. Stensgaard, J. Kjems, L. N. Kantorovich and F. Besenbacher, *Angew. Chem., Int. Ed.*, 2008, **47**, 9673.
28 L. Wang, H. Kong, C. Zhang, Q. Sun, L. Cai, Q. Tan, F. Besenbacher and W. Xu, *ACS Nano*, 2014, **8**, 11799.
29 S. Xu, M. Dong, E. Rauls, R. Otero, T. R. Linderoth and F. Besenbacher, *Nano Lett.*, 2006, **6**, 1434.
30 W. Xu, R. E. Kelly, H. Gersen, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich and F. Besenbacher, *Small*, 2009, **5**, 1952.
31 W. Xu, Q. Tan, M. Yu, Q. Sun, H. Kong, E. Lægsgaard, I. Stensgaard, J. Kjems, J. Wang, C. Wang and F. Besenbacher, *Chem. Commun.*, 2013, **49**, 7210.
32 W. Xu, J. Wang, M. F. Jacobsen, M. Mura, M. Yu, R. E. Kelly, Q. Meng, E. Lægsgaard, I. Stensgaard, T. R. Linderoth, J. Kjems, L. N. Kantorovich, K. V. Gothelf and F. Besenbacher, *Angew. Chem., Int. Ed.*, 2010, **49**, 9373.
33 C. Zhang, L. Wang, L. Xie, Y. Ding and W. Xu, *Chem. – Eur. J.*, 2017, **23**, 2356.
34 C. Zhang, L. Wang, L. Xie, H. Kong, Q. Tan, L. Cai, Q. Sun and W. Xu, *ChemPhysChem*, 2015, **16**, 2099.
35 C. Zhang, L. Xie, L. Wang, H. Kong, Q. Tan and W. Xu, *J. Am. Chem. Soc.*, 2015, **137**, 11795.
36 T. K. Shimizu, J. Jung, H. Imada and Y. Kim, *Angew. Chem., Int. Ed.*, 2014, **126**, 13949.
37 D. Skomski, S. Abb and S. L. Tait, *J. Am. Chem. Soc.*, 2012, **134**, 14165.
38 D. Skomski and S. L. Tait, *J. Phys. Chem. C*, 2013, **117**, 2959.
39 C. Wäckerlin, C. Lacovita, D. Chylarecka, P. Fesser, T. A. Jung and N. Ballav, *Chem. Commun.*, 2011, **47**, 9146.
40 Q. Fan, J. M. Gottfried and J. Zhu, *Acc. Chem. Res.*, 2015, **48**, 2484.
41 L. Grill, M. Dyer, L. Lafferentz, M. Persson, M. V. Peters and S. Hecht, *Nat. Nanotechnol.*, 2007, **2**, 687.
42 A. Rastgoo-Lahrood, J. Björk, W. M. Heckl and M. Lackinger, *Chem. Commun.*, 2015, **51**, 13301.