Molecular Recognition

DOI: 10.1002/anie.201003390

Supramolecular Porous Network Formed by Molecular Recognition between Chemically Modified Nucleobases Guanine and Cytosine**

Wei Xu, Jian-guo Wang, Mikkel F. Jacobsen, Manuela Mura, Miao Yu, Ross E. A. Kelly, Qiang-qiang Meng, Erik Lægsgaard, Ivan Stensgaard, Trolle R. Linderoth, Jørgen Kjems, Lev N. Kantorovich, Kurt V. Gothelf, and Flemming Besenbacher**

The involvement of surfaces in the origin of the first genetic molecules on Earth has long been suggested.[1] Prior to the emergence of nucleic acid polymerases in the prebiotic soup,[1,2] the self-assembly of primitive nucleobase building blocks may have relied on surface-mediated recognition events which catalyzed the formation of a covalent backbone in prototype oligonucleotides that subsequently may have functioned as templates in a primitive copying mechanism. This initial replication process may have been catalyzed by surfaces or chemical substances in solution—including RNA itself, as postulated in the RNA world hypothesis.[3,4] Today, the role and the relative importance of the basic, fundamental driving forces for nucleic acid replication such as base pairing, base stacking, and steric effects are still under intense debate. Watson–Crick hydrogen bonding has traditionally been thought to be a prerequisite for high-fidelity DNA replication.[5] However, recent studies on nucleobase analogues with the same size and shape as the natural ones but without relevant hydrogen-bonding groups have revealed that these analogues can recognize each other with high fidelity when incorporated into DNA sequences in vivo.[6,7] Watson–Crick hydrogen bonding thus seems not to be a requisite for the selectivity of base pairing in DNA replication. However, in the absence of polymerases in the prebiotic soup, Watson–Crick hydrogen bonding may have played a more crucial role in the molecular recognition between the nucleobase building blocks at surfaces and for further polymerization. In support of this postulation, molecular recognition between complementary bases, most likely driven by hydrogen bonding alone, has already been observed both at the liquid/solid (HOPG) interface and on the noble Au(111) surface under extreme ultrahigh vacuum (UHV) conditions.[8–10] These previous experiments were, however, conducted with nucleobases alone, and hence did not take the presence of deoxyribose into account. It is therefore of utmost importance to explore the role that Watson–Crick hydrogen bonding plays at surfaces in chemical structures that mimic nucleotides so as to address the fundamental question of how the polymerization of nucleotides may have started in the prebiotic soup in the absence of enzymes. The development of the scanning tunneling microscopy (STM) technique has advanced our understanding of supramolecular self-assembly systems on surfaces and has allowed intermolecular interactions to be explored at the submolecular scale.[11–15]

Herein we show by using a combination of high-resolution STM imaging and density functional theory (DFT) that sequential co-deposition of *N*-aryl-modified nucleobases[16,17] cytosine (C) and guanine (G) onto the Au(111) surface under UHV conditions results in the formation of highly ordered supramolecular porous networks, where Watson–Crick hydrogen bonding between chemically modified C and G molecules plays the primary role in their stabilization. As the N-arylation of the nucleobases has been performed on the nitrogen atom normally attached to the sugar moiety in DNA or RNA (Scheme 1), these *N*-aryl-modified nucleobases thus represent two-dimensional (2D) structural mimics of naturally occurring nucleotides. The current results outline a new route for directing the self-assembly of nucleobase-derived nanostructures at the surface. Furthermore, the observed

[*] Prof. W. Xu, Dr. M. F. Jacobsen, Dr. M. Yu, Prof. E. Lægsgaard, Prof. I. Stensgaard, Prof. T. R. Linderoth, Prof. J. Kjems, Prof. K. V. Gothelf, Prof. F. Besenbacher
Interdisciplinary Nanoscience Center (iNANO) and Center for DNA Nanotechnology (CDNA), Department of Physics and Astronomy, Department of Chemistry, and Department of Molecular Biology, Aarhus University
8000 Aarhus C (Denmark)
E-mail: fbe@inano.au.dk

Prof. W. Xu
College of Materials Science and Engineering
Tongji University
4800 Cao An Road, Shanghai 201804 (P.R. China)
E-mail: xuwei@tongji.edu.cn

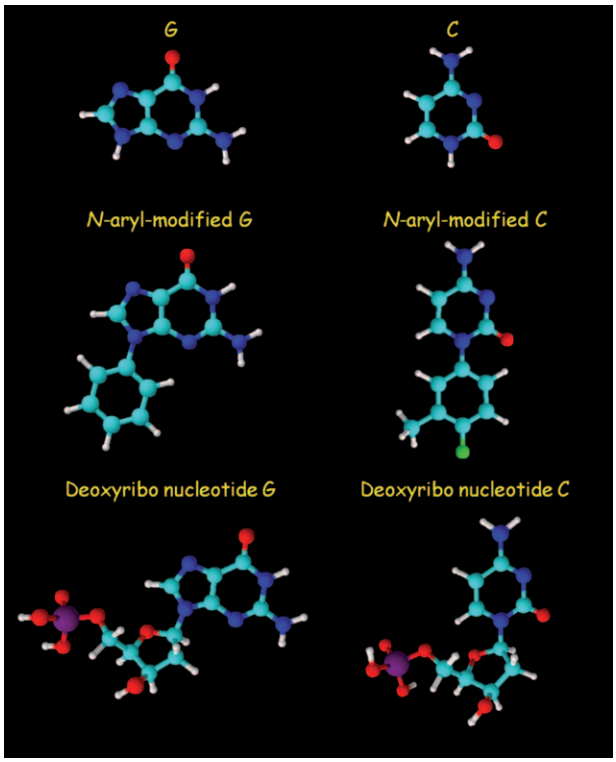
Prof. J.-g. Wang, Q.-q. Meng
Institute of Industrial Catalysis
College of Chemical Engineering and Materials Science
Zhejiang University of Technology
Hangzhou, 310032 (P.R. China)

Dr. M. Mura, Prof. L. N. Kantorovich
Department of Physics, School of Physical Sciences and Engineering, King's College London
The Strand, London, WC2R 2LS (United Kingdom)

Dr. R. E. A. Kelly
Department of Physics and Astronomy
University College London
Gower Street, London WC1E 6BT (United Kingdom)

[**] We acknowledge financial support from the Danish National Research Foundation, the Danish Ministry for Science, Technology, and Innovation through the iNANO Center, the Danish Research Councils, the Carlsberg Foundation, the Villum Kahn Rasmussen Foundation, the program for New Century Excellent Talents in University (NCET-10-0607) (W.X.), the EPSRC (grant GR/S97521/01), and the European Research Council (ERC) through an Advanced ERC grant (no 227430, VIN).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201003390.



Scheme 1. Chemical structures of G, C, *N*-aryl-modified G and C, as well as deoxyribo nucleotide G and C. O: red, N: blue, C: cyan, H: white, F: green, and P: purple.

supramolecular surface structures reveal how Watson–Crick hydrogen bonding in general may play a unique role for molecular recognition between complementary nucleobases.

Deposition of the *N*-aryl-modified nucleobase G led to a well-ordered chain structure, as deduced from the high-resolution STM image in Figure 1 a. Each individual molecule was imaged by STM as a triangle-like feature linked with a round feature that corresponds to the guanine moiety and the phenyl ring, respectively. The experimental unit cell ($a = (0.8 \pm 0.1)$ nm, $b = (2.2 \pm 0.1)$ nm, $\theta = (90° \pm 1)°$) is indicated in Figure 1 a. Several possible 2D structures formed by the modified G moiety were identified from the gas-phase DFT calculations (see the Supporting Information for details). The most stable structure (depicted in Figure 1 c), the formation energy of which is 1.34 eV per G pair, fits well with the STM image and the unit cell (the theoretical gas-phase values are $a = 0.82$ nm, $b = 2.2$ nm, $\theta = 90°$). It can be seen that both chiral forms of the modified G molecules are linked together by hydrogen bonding to form the zigzag chains. The interaction between the phenyl rings between these chains dominates the lateral assembly, thereby resulting in the formation of the observed 2D supramolecular nanostructure.

Deposition of the *N*-aryl-modified nucleobase C also led to a well-ordered chain structure (gure 1 b) The experimental unit cell ($a = (0.7 \pm 0.1)$ nm, $b = (2.0 \pm 0.1)$ nm, $\theta = (80° \pm 1)°$) fits well with the values obtained from the DFT calculations ($a = 0.74$ nm, $b = 1.9$ nm, $\theta = 80°$; Figure 1 d). The calculated energy of formation is 1.24 eV per C pair. Higher submolecular resolution by STM imaging of this structure was hampered because of the high mobility of these molecules on the Au(111) surface, even at low temperatures (ca. 120 K), and the relatively weak intermolecular interactions in the formed nanostructure; this was evidenced by the fact that the structure was easily disrupted by the STM tip during scanning. Similar to the case of the modified G structure discussed above, the modified C molecules in Figure 1 d are also linked together by hydrogen bonding to form a chain structure, where a weak lateral interchain binding is provided by the interacting phenyl rings and the methyl groups of the neighboring molecules to form the observed 2D supramolecular nanostructure.

Sequential co-deposition experiments of the two modified nucleobases were then performed. We took advantage of the high mobility of the modified C molecules on the surface to mix the two nucleobases as uniformly as possible: thus, we first deposited the modified C molecules at RT, which resulted in a 2D "fluid" on the surface which could not be imaged by STM at this temperature. Subsequently, the modified G molecules were deposited onto the partially C-covered Au surface (at RT), and thereafter the surface was cooled and imaged by STM at low temperatures. Very interestingly, co-deposition of these two complementary modified nucleobases resulted in a highly ordered porous network (with islands 30 nm across on average; Figure 2 a). This supramolecular network structure, which is distinctly different from the structures formed by the individual molecules described above, possesses a large unit cell with $a = (2.9 \pm 0.1)$ nm, $b = (2.9 \pm 0.1)$ nm, $\theta = (60 \pm 1)°$ (Figure 2 a).

A thorough and detailed analysis of the high-resolution STM images was carried out to gain further insight into this supramolecular network structure formed from the mixture of complementary modified G and C molecules at the atomic scale. From the close-up STM image depicted in Figure 2 b we can distinguish two different kinds of molecules: some appear as triangle-linked-ball protrusions, which are assigned, as above, to the modified G molecules, while the other molecules appear as elongated protrusions with subtle subfeatures, which are thus assigned as the modified C molecules. It thus appears that during the co-deposition experiments the modified G molecules get incorporated into the preexisting fluid of the modified C molecules, thereby resulting most likely in the formation of G-C dimers in the supramolecular porous network. The identification of the modified C molecules in the binary mixture phase in the STM images implies that the intermolecular interaction between the modified G and C molecules is stronger than the intermolecular interaction between the modified C molecules alone, and enables us to obtain the submolecular-resolution images on both modified C and G molecules in the binary mixture phase. From the STM results we tentatively conclude that such G-C dimers are periodically distributed within the supramolecular network as the elementary building blocks.

To verify this hypothesis and work out the corresponding atomic-scale model for the mixture phase, DFT calculations were performed on several models by using the G-C Watson–Crick base pair as the starting point, since such a base pair has

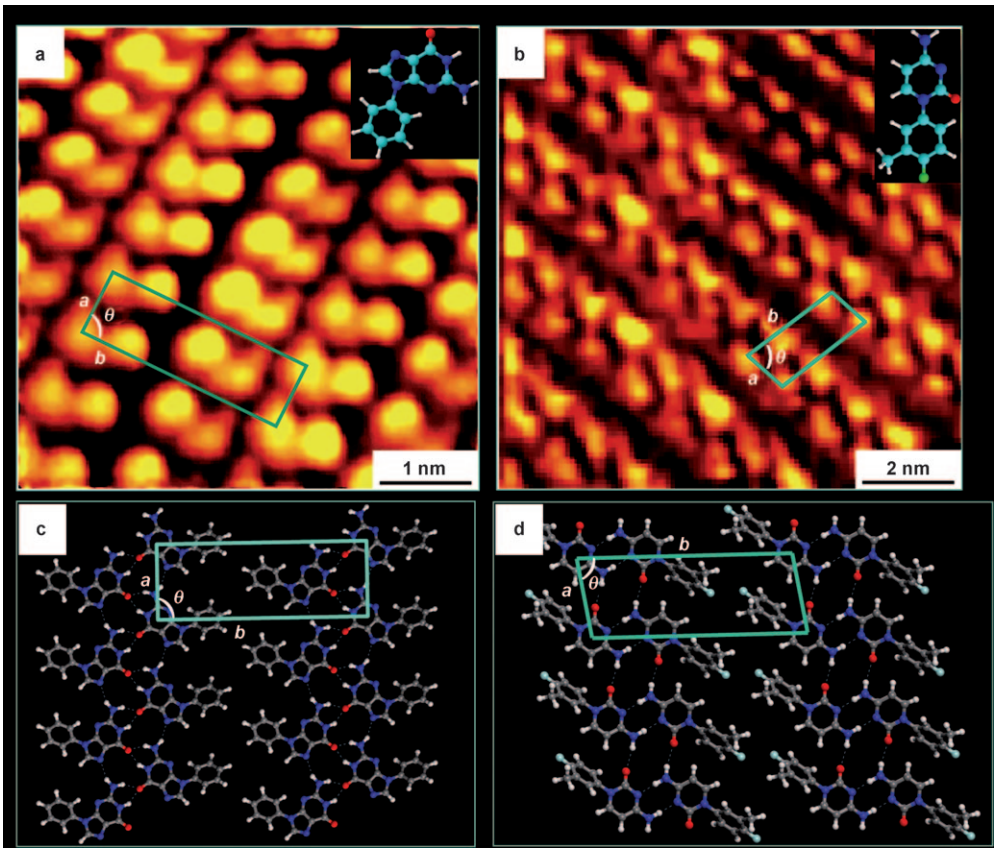


Figure 1. High-resolution STM images of self-assembled surface nanostructures formed by *N*-aryl-modified nucleobases G (a) and C (b). The insets represent the corresponding chemical structures. Scanning conditions: $I_t = -0.7$ nA, $V_t = -1051$ mV for (a) and $I_t = -0.72$ nA, $V_t = -1250$ mV for (b). The DFT calculated molecular models are presented in (c) and (d) for *N*-aryl-modified G and C, respectively. The dotted lines indicate intermolecular hydrogen bonds. The unit cells are also indicated in each case.

the highest stability among all the nucleobase pairs. We have also calculated the STM images of various G and C pairs on the two-layer Au(111) surface (see the Supporting Information for details).

From a detailed comparison of all the theoretically calculated structures as well as the simulated STM image of the G-C Watson–Crick base pair with the high-resolution STM images we obtained one unique model which seems to fit well with the experimental data; the superimposition of this model on the STM image is depicted in Figure 2b (the unit cell values from the DFT calculations are $a = 2.8$ nm, $b = 2.8$ nm, $\theta = 60°$). In this model the G-C Watson–Crick base pairs (highlighted by green ellipsoids) are distributed uniformly within the supramolecular mixture phase. We thus conclude that the high stability and structural ordering of this supramolecular porous network are due to the formation of highly stable G-C Watson–Crick base pairs triggered by the formation of triple hydrogen bonds. Furthermore, weak hydrogen bonds and possible van der Waals interactions with the aryl substituent may help in the secondary ordering between the G-C Watson–Crick base pairs. The calculated formation energies of the G-C Watson–Crick base pair (Figure 2c), the larger building block (Figure 2d), and the supramolecular porous network (Figure 2e), all calculated in the gas phase, are 1.29, 1.58, and 1.62 eV, respectively, per single G-C pair. Hence, we propose the following hierarchical self-assembly scenario (Figure 2c–e), where three fundamental steps can be identified: 1) the modified C and G molecules first recognize each other to form the Watson–Crick base pairs (Figure 2c); 2) three G-C Watson–Crick base pairs are linked together through weak hydrogen bonds (CH···O and CH···N) to form the larger building block for the subsequent assembly process (Figure 2d); 3) these larger building blocks are further connected together through weak hydrogen bonds (CH···F) and possible van der Waals interactions, thereby resulting in the final supramolecular porous network (Figure 2e). Importantly, we find that the recognition between the modified C and G molecules is the key to the formation of the final porous network.

Adsorption of such chemically modified nucleobases on a noble Au(111) surface under well-controlled UHV conditions constitutes an interesting model system that may give unique insight into the fundamental driving forces behind molecular recognition between complementary nucleobases. The Au(111) surface served as a template to ensure that the nucleobases adopt a flat adsorption geometry[8,18–22] that mimics the base-stacking situation in natural double-stranded nucleic acids (see the Supporting Information), and the well-controlled ultraclean UHV conditions provide an idealized environment without water, salts, metal ions, etc. The main determinants for intermolecular interactions between the nucleobases under these extreme conditions are limited to hydrogen bonding and possible van der Waals interactions between the aryl groups. It is reasonable to expect that higher efficiency and accuracy in the molecular recognition process exist between the biomimetically modified nucleobases compared to the nucleobases C and G alone, because fewer competitive hydrogen-bonded configurations are available for the modified nucleobases. This is indeed confirmed by our previous experiments[8] performed with the unmodified bases C and G, where we found that Watson–Crick G-C pairs were only randomly embedded into a C filament and characteristic fivefold ring structures without higher ordering.

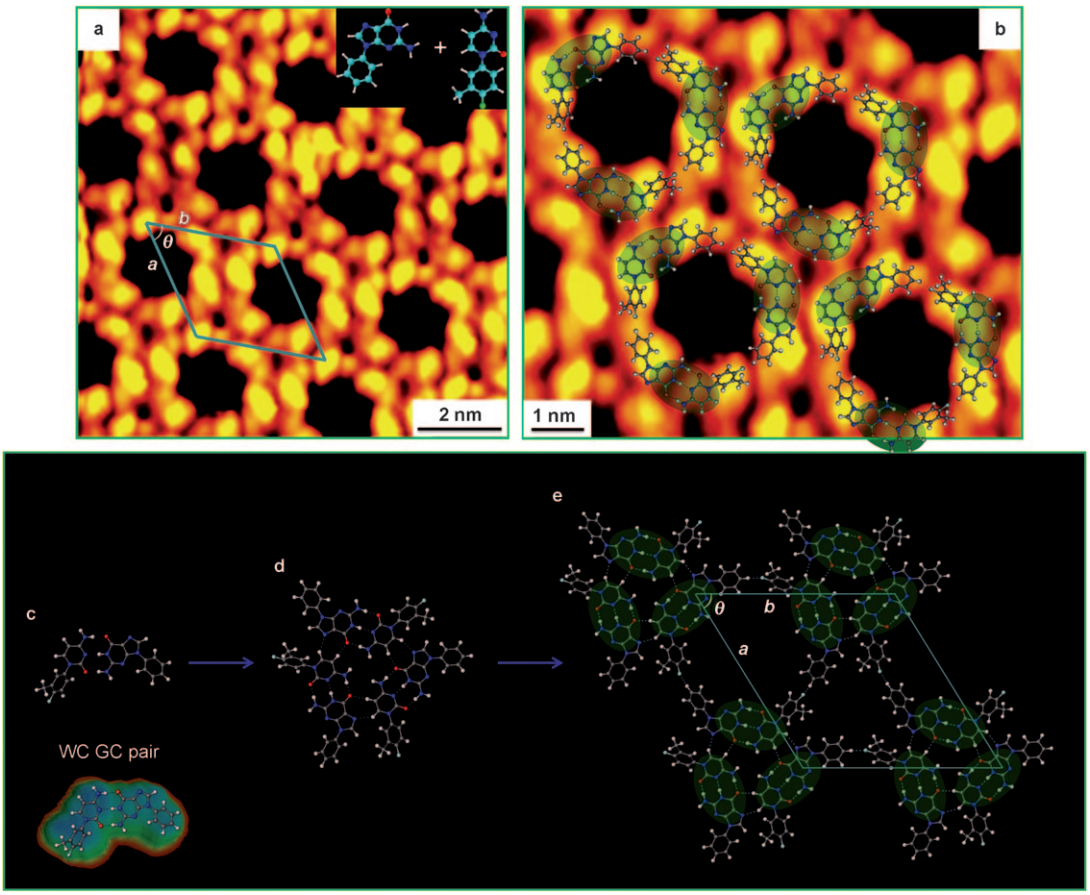


Figure 2. a) High-resolution STM image of a self-assembled supramolecular porous network formed by sequential co-deposition of *N*-aryl-modified nucleobases C and G. Scanning conditions: $I_t = -1.7$ nA, $V_t = -2500$ mV. b) DFT-calculated molecular model superimposed on the close-up STM image. The Watson–Crick GC pairs are highlighted by green ellipsoids. c)–e) show the theoretical models for the proposed hierarchical self-assembly of the porous network. A simulated STM image of the G-C Watson–Crick base pair on the Au(111) surface is also presented in (c).

In summary, we have demonstrated that Watson–Crick hydrogen bonding can indeed play a vital role in driving natural molecular recognition between complementary nucleobases on a noble gold surface, even under well-controlled UHV conditions. The N-arylated nucleobases may be viewed as 2D mimics of natural nucleotides, and their efficient Watson–Crick hydrogen-bonding pattern on surfaces gives strong support to the idea that the self-assembly of nucleobase building blocks may have served as a starting point for the formation of nucleic acids in the prebiotic soup prior to the emergence of polymerases. Further experiments on varying the geometrical configurations and physicochemical properties of the artificial groups linked to the nucleobases are under investigation, with the aim of exploring the role of steric effects in the molecular recognition process at a single nucleobase level.

Experimental Section

The experiments were performed in a UHV chamber equipped with a variable-temperature, high-stability, Aarhus STM.[23] The C and G derivatives were thermally sublimated onto Au(111) at 390 K and 430 K, respectively, and the subsequent STM images were recorded at 100–160 K. See the Supporting Information for further details.

The theoretical predictions of the structures were carried out in the gas phase by using first-principles DFT calculations based on periodic boundary conditions, a localized basis set, and the method of pseudopotentials. Separate calculations on a single molecule adsorbed on the Au(111) surface which took into account the van der Waals interaction between them revealed strong binding with negligible lateral dependence, thus rendering the molecules very mobile on the surface at RT and, at the same time, justifying the gas-phase modeling of supramolecular assemblies. The STM images were calculated by using the scattering method. The Supporting Information contains details of all these calculations.

Received: June 4, 2010
Revised: October 11, 2010
Published online: November 9, 2010

Keywords: density functional calculations · hydrogen bonds · molecular recognition · nucleobases · scanning probe microscopy

[1] S. J. Sowerby, W. M. Heckl, *Origins Life Evol. Biosphere* **1998**, *28*, 283.
[2] J. L. Bada, A. Lazcano, *Science* **2002**, *296*, 1982.

[3] L. E. Orgel, *Nature* **1992**, *358*, 203.
[4] G. F. Joyce, *Nature* **1989**, *338*, 217.
[5] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.
[6] E. T. Kool, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1.
[7] J. C. Delaney, P. T. Henderson, S. A. Helquist, J. C. Morales, J. M. Essigmann, E. T. Kool, Proc. *Natl. Acad. Sci. USA* **2003**, *100*, 4469.
[8] R. Otero, W. Xu, M. Lukas, R. E. A. Kelly, E. Lægsgaard, I. Stensgaard, J. Kjems, L. N. Kantorovich, F. Besenbacher, *Angew. Chem.* **2008**, *120*, 9819; *Angew. Chem. Int. Ed.* **2008**, *47*, 9673.
[9] W. Mamdouh, M. Dong, S. Xu, E. Rauls, F. Besenbacher, *J. Am. Chem. Soc.* **2006**, *128*, 13305.
[10] S. Xu, M. Dong, E. Rauls, R. Otero, T. R. Linderoth, F. Besenbacher, *Nano Lett.* **2006**, *6*, 1434.
[11] J. V. Barth, G. Costantini, K. Kern, *Nature* **2005**, *437*, 671.
[12] R. Otero, F. Rosei, F. Besenbacher, *Annu. Rev. Phys. Chem.* **2006**, *57*, 497.
[13] L. J. Wan, *Acc. Chem. Res.* **2006**, *39*, 334.
[14] S. De Feyter, F. C. De Schryver, *Chem. Soc. Rev.* **2003**, *32*, 139.
[15] A. Ciesielski, S. Lena, S. Masiero, G. P. Spada, P. Samorì, *Angew. Chem.* **2010**, *122*, 2007; *Angew. Chem. Int. Ed.* **2010**, *49*, 1963.
[16] M. F. Jacobsen, C. S. Andersen, M. M. Knudsen, K. V. Gothelf, *Org. Lett.* **2007**, *9*, 2851.
[17] M. F. Jacobsen, M. M. Knudsen, K. V. Gothelf, *J. Org. Chem.* **2006**, *71*, 9183.
[18] R. Otero, M. Schöck, L. M. Molina, E. Lægsgaard, I. Stensgaard, B. Hammer, F. Besenbacher, *Angew. Chem.* **2005**, *117*, 2310; *Angew. Chem. Int. Ed.* **2005**, *44*, 2270.
[19] R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Science* **2008**, *319*, 312.
[20] W. Xu, R. E. A. Kelly, R. Otero, M. Schöck, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2007**, *3*, 2011.
[21] R. E. A. Kelly, W. Xu, M. Lukas, R. Otero, M. Mura, Y.-J. Lee, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2008**, *4*, 1494.
[22] W. Xu, R. E. A. Kelly, H. Gersen, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2009**, *5*, 1952.
[23] E. Laegsgaard, L. Österlund, P. Thostrup, P. B. Rasmussen, I. Stensgaard, F. Besenbacher, *Rev. Sci. Instrum.* **2001**, *72*, 3537.

