

Understanding the disorder of the DNA base cytosine on the Au(111) surface

Ross E. A. Kelly,^{1,2,a)} Maya Lukas,^{3,4} Lev N. Kantorovich,¹ Roberto Otero,^{3,5} Wei Xu,³ Manuela Mura,¹ Erik Lægsgaard,³ Ivan Stensgaard,³ and Flemming Besenbacher³

¹Department of Physics, School of Physical Sciences and Engineering, King's College London, Strand, London WC2R 2LS, United Kingdom

²Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom

³Interdisciplinary Nanoscience Center (iNANO), University of Aarhus, 8000 Aarhus C, Denmark, Center for DNA Nanotechnology (CDNA), University of Aarhus, 8000 Aarhus C, Denmark, and Department of Physics and Astronomy, University of Aarhus, 8000 Aarhus C, Denmark

⁴Institute for Nanotechnology, Forschungszentrum Karlsruhe, Postfach 3640, D-76021 Karlsruhe, Germany

⁵Departamento de Física de la Materia Condensada, Universidad Autónoma de Madrid, 28049 Madrid, Spain

(Received 15 May 2008; accepted 24 September 2008; published online 14 November 2008)

Using ultrahigh vacuum scanning tunneling microscopy (STM) and *ab initio* density functional theory, we have investigated in detail structures formed by cytosine on the Au(111) surface in clean ultrahigh vacuum conditions. In spite of the fact that the ground state of this DNA base on the surface is shown to be an ordered arrangement of cytosine one-dimensional branches (filaments), this structure has never been observed in our STM experiments. Instead, disordered structures are observed, which can be explained by only a few elementary structural motifs: filaments, five- and sixfold rings, which randomly interconnect with each other forming bent chains, T junctions, and nanocages. The latter may have trapped smaller structures inside. The formation of such an unusual assembly is explained by simple kinetic arguments as a liquid-glass transition. © 2008 American Institute of Physics. [DOI: 10.1063/1.3001585]

I. INTRODUCTION

In recent years, the self-assembly of DNA base molecules on solid surfaces (self-assembled superstructures) has attracted much attention, both experimentally and theoretically,^{1–32} due to possible applications in the emerging area of nanotechnology.^{33–38} It has also been suggested¹¹ that studying DNA base molecules is necessary to understand the formation of complex biological molecules responsible for the origin of life forms on our planet billions of years ago. In particular, the DNA base cytosine has been suggested to play a key role in the origin of homochirality of biomolecules.³⁹

Scanning tunneling microscopy (STM) has been indispensable in studying DNA base molecules on metal surfaces, providing useful information on molecular arrangements. As was shown previously,^{16,17,28} the Au(111) surface serves as an ideal substrate for studying DNA base self-assemblies with STM. For instance, it was shown¹⁶ that guanine molecules are able to self-assemble into ordered two-dimensional (2D) assemblies on this surface. Similar findings have been obtained for adenine on Au(111),²⁸ where it was found that the perturbation induced by the surface on the hydrogen bonding between molecules is very small. In comparison, thymine on the Au(111) surface forms chains at low coverage, but forms monolayers when the coverage is increased to a certain level due to the van der Waals interaction between the molecular hydrogen bonded chains.¹⁷

Many previous studies of cytosine (C) assemblies on the Au(111) surface have been performed at the liquid-solid interface.^{4,6,24,40,41} Tao *et al.* reported that C in solution forms a monolayer on the Au(111) surface spontaneously.⁴ Wandlowski *et al.* found that C molecules in solution adsorb onto the Au(111) surface, forming both an ordered and disordered phases.⁶ Other investigations of C in solution adsorbed on the Au(111) surface also show 2D growth.^{24,41} However, since all the observed studies in solution depend on the applied voltage as well as on the pH of the solution used, these assemblies are expected to be strongly affected by the solution itself and their analysis is much more complex.

Cytosine deposited under well-controlled clean ultrahigh vacuum (UHV) conditions was observed to form one-dimensional chains (or filaments) on a Cu(111) surface.^{5,9,14} However, the molecular resolution in the images was poor, and the origin of the filamentary structure was not clear. It has been suggested that small C clusters are the basic units of the observed structures.^{5,9} However, this proposal was not pursued further in later studies where a molecular zigzag chain was proposed.¹⁴ The chemical inertness of the Cu(111) surface to the C molecules allows for their easy diffusion, facilitating formation of superstructures via hydrogen bonding between C monomers.⁵ Note that only 1D C superstructures (chains or filaments) have been observed so far under UHV conditions, the only exception being a regular 2D structure observed on a corrugated Cu(110) surface⁴² in which the molecules were found to be standing upright because of a strong interaction with the surface.

^{a)}Electronic mail: ross.kelly@ucl.ac.uk

However, it is not trivial, in general, to determine the detailed atomic-scale structure from recorded STM images. In the cases of the periodic structures, it is often only possible to deduce the lattice vectors and sometimes the space symmetry of the observed superstructures. Unfortunately, in most cases, small molecules appear as single bright protrusions (of some nearly round shape) in STM images, so that their orientation usually cannot be deduced. Therefore, theoretical calculations are necessary to investigate the atomic scale structure of many small molecular assemblies.

Previous theoretical investigations mainly relied on calculations of *selected* base pairs, which, after being ordered in terms of their stabilization energies, were used to construct models of the observed structures.^{14,15,20,26} However, to our knowledge, there have been no previous detailed theoretical calculations of C superstructures. A model for a single C chain was proposed in Ref. 14, using three homopair possibilities obtained from a semiempirical calculation. However, this model was not verified in actual simulations.

Several points are worth considering for suggesting appropriate molecular models. First, the energetically most favorable superstructures may not necessarily be formed exclusively by the most stable pairs. For example, the guanine tetrad structure reported recently¹⁶ is not based entirely on the most stable centrosymmetric homopair found in Ref. 43. Second, even in the case of a weak molecule-surface interaction, one cannot completely neglect the effect of the surface on the formed structures.^{18,28} Third, as it was shown in the case of adenine, many structures may exist with similar periodicity patterns.^{18,27,28} Therefore, in order to characterize the experimentally observed structures, a *systematic* approach, similar to that developed for adenine,^{18,27,28} which is able to predict *all* possible structures, is needed, so that the correct interpretation of the STM images can be made. In doing so, all known homopairings⁴³ should be used with the proper account of the possible constraints imposed by the substrate. The most “probable” structures are then considered in detail using *ab initio* methods. Note that semiempirical approaches should be used with care when considering DNA base assemblies as both the geometry and the energetics they predict are unreliable.^{43–46} This approach that combines STM experimental images with a systematic theoretical analysis of all possible structures enables us to make a substantial step forward from “guessing” to “deriving” the observed molecular structures.

In this paper, we provide a detailed account of the structures that C molecules form on the Au(111) surface. In Sec. II, results of our STM observations of C networks on Au(111) are reported exhibiting a high degree of disorder and containing two basic building blocks: zigzag chains and five- and sixfold rings. These interconnect with each other yielding a complicated random network. Some of the molecules get trapped inside the network nanocages (cavities) and cannot escape. In Sec. III, we show, by using an approach based on a systematic generation of all possible structures with the subsequent *ab initio* density functional theory (DFT) treatment, that all basic building blocks are indeed stable structures. Furthermore, the calculations enable us to understand the origin of the observed disorder including the

nature of all possible interconnections between the elementary building blocks. In Sec. IV, we describe the kinetics of formation of the observed random network. Finally, conclusions are drawn in Sec. V. A brief account of this study has been recently published.¹

II. EXPERIMENTAL RESULTS

A. Experimental method

All STM experiments were performed in a UHV chamber (base pressure 1×10^{-10} Torr) equipped with the variable-temperature, fast-scanning Aarhus STM,^{47,48} a home-built molecular evaporator, and standard facilities for sample preparation. Au(111) single crystals were cleaned by repeated Ar ion sputtering at 3 kV and consecutive annealing to 730–780 K, resulting in a clean, flat, well-ordered ($22 \times \sqrt{3}$) herringbone reconstructed Au(111) surface.⁴⁹ Molecules (C powder, purity 99%, from Sigma Aldrich) were evaporated from the glass crucible of the home-built molecular evaporator. The evaporator, which is separated from the main chamber by a gate valve, was pumped and baked after loading, and the glass crucibles were degassed at temperatures slightly below deposition temperature for several days prior to deposition. Directly after cleaning of the Au(111) surface, the evaporator (at temperatures of 400–420 K) was inserted into the main chamber, and then the molecules were deposited onto the Au sample held at room temperature (RT). The sample was then immediately transferred to the pre-cooled STM, and the images were recorded at $T = 100$ – 150 K in the constant current mode [$I = 0.4$ – 0.7 nA, $U = \pm(0.8$ – $3.0)$ V]. Both negative and positive voltages were used to probe highest occupied molecular orbital and lowest unoccupied molecular orbital resonances of the molecule which are separated in energy by just over 4 eV. Note that the constant current mode is advantageous in providing better resolution of the molecule on the surface.

B. Observations

Figure 1 shows a STM image of a typical sample at low [(a) and (b)], medium (c) and high (d) C coverages. A loose disordered network of *zigzag lines* and *rings* is clearly visible at the low and medium coverages. In strong contrast to the deposition of guanine¹⁶ and adenine,²⁸ which all establish 2D *ordered* structures on the Au(111), C is the DNA base that grows only into a disordered network of 1D branches (filaments). It should be noted that thymine forms both ordered and disordered structures on the same surface.¹⁷ At coverages close to a monolayer (ML), the network becomes more dense, but the branched structure consisting of zigzag filaments is still visible, Fig. 1(d). We never observed any well-ordered 2D structure on the Au(111) surface, even at the highest coverage.

The bright circles making up this network fit the expected apparent size of a single C molecule. We thus conclude that the branches consist of 1D zigzag lines of molecules, which interconnect in a random fashion. These branches are often terminated by a ringlike structure that contains five molecules. However, interconnections between

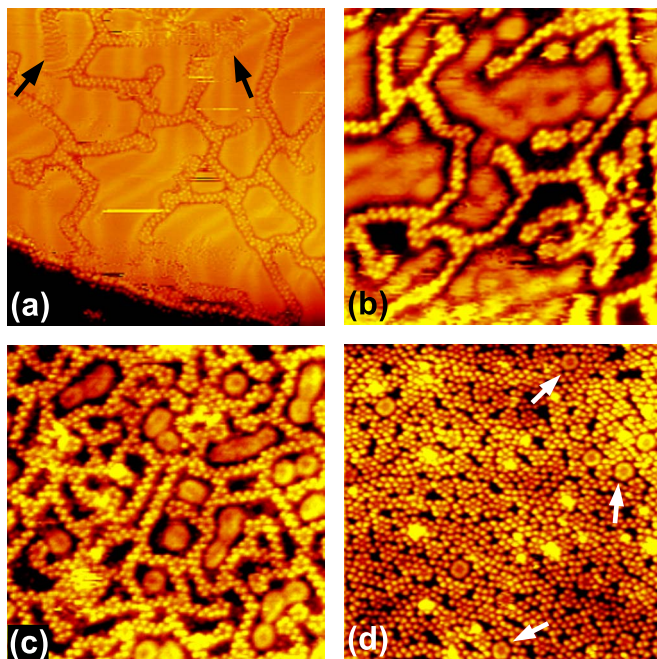


FIG. 1. (Color online) Development of the cytosine structure with increasing coverage. (a) The herringbone reconstruction is not lifted and the zigzag branches are able to move in a whiplike manner (indicated by the black arrows). (b) Confined areas contain a number of blurs, which consist of moving C clusters; branches are usually terminated by fivefold rings, and junctions between interconnecting branches are often established by sixfold rings. (c) With increasing coverage, the number of mobile C molecules in blurred spots decreases, while the number of sixfold rings increases. (d) At high coverages, the structure is still disordered, and the basic units such as zigzag filaments and rings are clearly observed; blurred formations confined in very small nanocages are less mobile and can sometimes be identified as five- or sixfold rings (indicated by white arrows). C coverage, image size, I , V : (a) 0.22, 450 Å, 0.4 nA, -2.9 V; (b) 0.22, 300 Å, 0.3 nA, 1.3 V; (c) 0.54, 300 Å, 0.3 nA, 1.2 V; (d) 0.75, 300 Å, 0.2 nA, -1.3 V. Note that no change in the observed structures could be found by changing the voltage polarity.

branches are very often established by sixfold rings. These three elementary structures (zigzag branches and five- and sixfold rings) are observed at all coverages.

Although, at first sight, most branches seem to follow three main directions with 120° between them, at a closer look, the branches are found to extend in all directions and can even be seen to bend. Therefore, there seems to be no registry with the underlying Au(111) surface. Furthermore, as in the case of the other bases, the C coverage does not lift the herringbone reconstruction of the underlying gold [Fig. 1(a)], indicating a very weak interaction between the bases and the substrate. This finding agrees with our previous experimental and theoretical (based on DFT calculations) analyses for adenine²⁸ and guanine¹⁶ on the same surface. Recent dynamics calculations of DNA bases on the Au(111) surface⁵⁰ indicate that the interaction of C with the gold is the weakest amongst all the bases. The high mobility of C on gold makes the imaging at low coverages extremely difficult or even impossible, and images at low coverage contain a high content of noise in most cases. Moreover, it can be observed that the structures are quite mobile on the surface: branches that are only connected to the network at one side move in a whiplike manner at temperatures as low as 100 K [see diffuse shapes marked by an arrow in Fig. 1(a)].

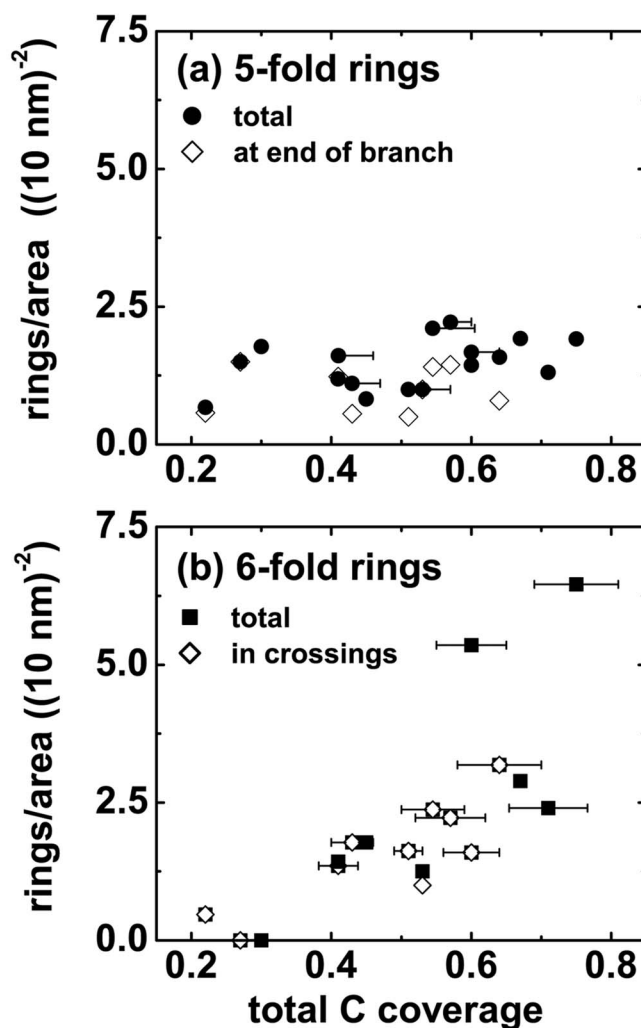


FIG. 2. Counts of five- (a) and sixfold (b) rings per area vs total cytosine coverage on the surface. 0.75 is the highest coverage observed, which corresponds to 1 ML. The number of fivefold rings stays nearly constant (within the error bars), while the number of sixfold rings clearly grows with increasing coverage. Closed symbols are total counts, while open symbols are either fivefold rings found as termination of branches (a) or sixfold rings found in branch junctions. The error bars correspond to the errors due to: (a) blurs and (b) cutoff heights for filaments (see text).

Areas that are completely confined by branches are often found to contain blurred bright features which we believe are small clusters (e.g., rings) of C molecules trapped inside these *nanocages*. If we open a nanocage by STM manipulation (with the tip), the blurred feature can be found to diffuse in and out from one area to the other.¹ Bigger cages do not contain big blurred features; instead, a number of smaller blurred features are visible, see Figs. 1(b) and 1(c). At larger coverages, the structure of single blurs in very narrow nanocages can sometimes be identified as a moving ring [Fig. 1(d)].

Having identified the branches, as well as the five- and sixfold rings as the characteristic elements, or *building blocks*, of the disordered network, we investigated how the network develops with increasing coverage. Figure 2 shows the number of five- and sixfold rings per area as a function of the total C coverage. The latter is defined as the area covered by molecules, with respect to the total available area on the

surface. By this definition, the coverage never reaches 1.0 due to spaces between the molecules. The highest coverage we observed for our disordered structure of C was 0.75, which we define as corresponding to 1 ML. For comparison, the relative area that is covered by 1 ML of close-packed, identical circular-shaped molecules of the same radius is 0.9.

The total C coverage estimated from an image can vary by changing the cutoff height, which defines the edge (the rim) of the molecule in the STM image. Due to the filamentary structure, a slight change in this cutoff may result in a change of the total C coverage by 0.02–0.06. This error increases with coverage [see error bars in Fig. 2(b)]. Furthermore, the coverage by molecules, which are contained in blurred features, gives rise to an additional error, see Fig. 2(a). At coverages at which clearly separated round blurs could be identified, we assumed that a blur is most likely to be a six- or fivefold ring (see Sec. III C 6). This error strongly decreases with increasing coverage. Below a coverage of 0.4, it is not possible to estimate the number of C in large blurred areas. At coverages below 0.2, no reliable quantitative estimations of the structures can be done due to the high mobility of the molecules on the surface.

Though the error bars are quite large, some clear trends can be found in Fig. 2. At the lowest coverages where imaging is difficult, zigzag filaments and fivefold rings terminating the branches are already found, see Fig. 1. Sixfold rings are very scarcely found, and their density is around 10^{-3} nm^{-2} . By increasing the C coverage, the number of fivefold rings fluctuates around the same number of $0.015 \pm 0.005 \text{ nm}^{-2}$. In contrast, the number of sixfold rings clearly increases with the coverage.

Figure 1 also confirms that at not too large coverages, fivefold rings are mainly found as terminations of branches, while sixfold rings are mainly found as *vertices* connecting three branches. Only in very few cases have we found structures that can unambiguously be identified as rings and that are not connected to branches, see Fig. 1(b). These are probably fixed at surface imperfections that, though rare, exist on the gold surface. Note that at high coverages when the branches close up, it is often difficult to distinguish whether a ring is either connected to the end of one or several branches. Except for step edges, no structure could clearly be identified as a single molecule or pair.

III. THEORETICAL ANALYSIS

In this section, it is described in detail how the cytosine molecules can join to one another to form a variety of structures, including dimers, chains, and ring structures. Furthermore, we describe how these species may join to one another, which allows for the observed disorder on the Au(111) surface (see Sec. II) to be explained.

A. Cytosine dimers

The main building block of any C superstructure is formed by a cytosine dimer. A systematic analysis of all possible C dimers calculated using an *ab initio* DFT method was recently presented in Ref. 43. Three types of binding sites have been identified as shown schematically in Fig. 3: (I)

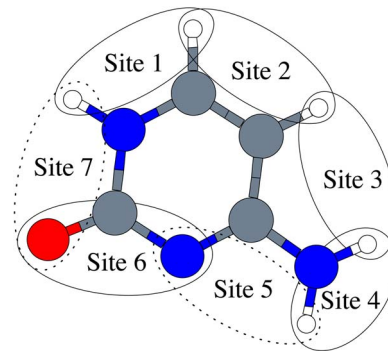


FIG. 3. (Color online) The cytosine molecule in configuration C. The chiral counterpart, the configuration \bar{C} , is obtained from C upon flipping in the molecular plane. The seven binding sites which can participate in forming a double hydrogen bond between cytosine homopairs are explicitly indicated. Site 8 which contains 3 atoms is the combination of sites 5 and 6.

with two hydrogens (sites 1–4), (II) with two acceptors (site 6), and (III) with one hydrogen and one acceptor (sites 5 and 7). The sites of type I can bind to site 6 (of type II), while the sites of type III can bind only to each other. Note that, for simplicity, we do not introduce here three-atom sites as these are essentially combinations of two adjacent two-atom sites.⁴³ For example, site 8 (which we use later) contains the combination of atoms from sites 5 (the hydrogen) and 6 (oxygen), and can join with itself to create C_8C_8 . This is essentially a type III-III bond involving two atoms from each molecule.

By combining two molecules in all possible ways and performing DFT calculations, 11 dimers were found in total, with the binding energies ranging between -0.2 and -1.0 eV .⁴³ In what follows, any dimer formed by binding site n of one molecule and site m of the other will be indicated as C_nC_m or $C_n\bar{C}_m$, depending on the chirality of the molecules C (shown in Fig. 3) and \bar{C} , the latter obtained from C by flipping the molecule C in the molecular plane. There are three chiral centrosymmetric pairs (C_5C_5 , C_7C_7 , and C_8C_8), each having C_2 symmetry. Together with the pair $C_5\bar{C}_7$, they form the first four most stable pairs, the pairs involving a combination of the sites 5 and 7 being the most stable.

B. *Ab initio* method

Similarly to a work on C dimers⁴³ and other previous studies on DNA base pairs,^{45,46} we use the *ab initio* DFT SIESTA (Refs. 51 and 52) method, which utilizes a localized numerical atomic orbital basis set, norm-conserving pseudopotentials, and periodic boundary conditions. In all calculations, we used the DZP (double zeta plus polarization orbitals) basis set with the energy cutoff of 10 meV, and the Perdew-Becke-Ernzerhof⁵³ density functional for the exchange-correlation energy, which was previously found to be adequate for DNA bases.^{43,45,46,54}

Atomic relaxation was performed until the forces on the atoms were less than $0.05 \text{ eV}/\text{\AA}$. The large cell sizes allowed us to use only a single (gamma) k point in our calculations. Energetics of relaxed structures were analyzed using^{43,45,46}

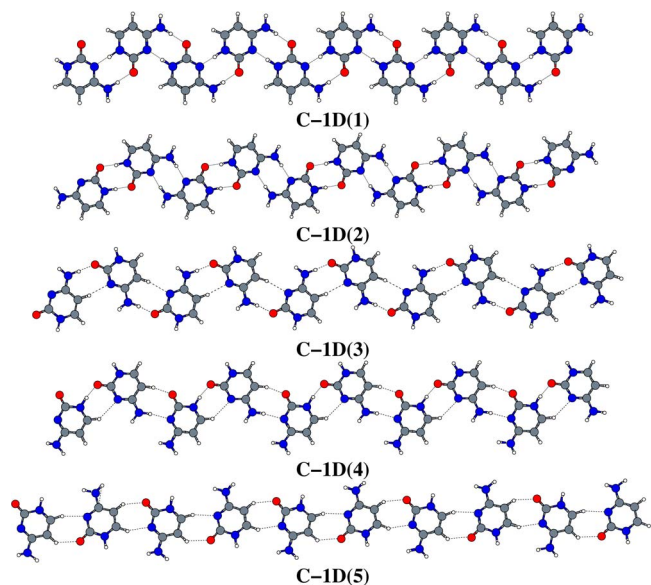


FIG. 4. (Color online) Relaxed one-dimensional cytosine filaments, denoted C-1D(N), where N in brackets indicates the order of stability according to Table I.

(i) the stabilization energy, E_{stab} , defined as the cell energy minus the sum of total energies of individually relaxed monomers; (ii) the interaction energy, E_{int} , defined as the cell energy minus the sum of the total energies of C monomers in the geometry of the combined system; and (iii) the deformation energy, $E_{\text{def}} = E_{\text{stab}} - E_{\text{int}}$, which characterizes the sum of energy losses of the monomers when forming the structure. When calculating the stabilization energies, the basis set superposition error (BSSE) correction⁵⁵ E_{BSSE} was applied in each case to account for the incompleteness of the basis set.

C. Understanding disorder

The C molecule, Fig. 3, has a special arrangement of binding sites which is different from that of guanine and adenine. Indeed, for the latter two molecules, donors and acceptors alternate around the molecules perimeter; C, in contrast, has only two acceptors (both contained in site 6) and five donors (contained in sites 1–4) which are not alternating but are adjacent to each other. This special arrangement of binding sites restricts significantly the ability of the molecules to form large superstructures. In particular, as we shall argue below, C molecules cannot form *ordered* 2D monolayers by hydrogen bonding only. At the same time, one-dimensional (1D) structures can easily be formed facilitating clustering of C molecules in one direction.

In order to characterize the structures observed in our images in Fig. 1, we first note, as suggested by observations of Sec. II B and our DFT calculations, that the interaction of C molecules with the surface is of less importance than their interaction with each other. Indeed, only a small energy gain of -0.1 eV was found for both planar geometries (C and \bar{C}) of the molecule on the Au(111) surface (the latter was modeled by a 5×5 slab of six layers), and the adsorption energy practically does not change when moving the molecule laterally across the surface. Note that the inclusion of the van der Waals interaction between the molecule and the surface,

which is missing in our DFT calculations (following the analytical formulas developed in Ref. 56), does indeed increase the adsorption energy. However, this energy still shows a negligible corrugation across the surface. Therefore, similarly to our previous studies on adenine superstructures,^{18,27,28} we can systematically construct various C assemblies by considering all possible structures in the *gas phase*, i.e., without taking the surface explicitly into account.

1. One-dimensional filament structures

To construct all possible filaments, we can connect several molecules in a chainlike structure and then repeat it periodically along the same direction. For simplicity, we assumed the periodicity of two molecules in the cell (cf. Ref. 14). Hence, each unit cell may be constructed out of a single dimer, and there are 11 different dimers possible. However, maintaining periodicity imposes some restrictions on the allowed dimers since the second molecule in the cell should connect to the image of the first molecule in the next cell. Only eight of the dimers can provide an extra binding site for the periodic filament structure, i.e., only eight filaments can be constructed with two molecules in the unit cell.

The eight possible filaments were fully relaxed using the *ab initio* SIESTA method described in Sec. III B. Only five possibilities were found, and these are shown in Fig. 4. From the three failed possibilities, one (made of by dimers C_7C_7 and C_8C_8) relaxed into structure C-1D(2), while the other two possibilities (which are periodic repeats of either C_2C_6 or C_6C_4 pairs) relaxed into nonchainlike structures. The structures of the five stable filaments alongside their relevant energies are given in Table I. The energy E_{pred} is an estimate of the stabilization energy calculated just as a sum of *ideal* dimer energies.

The second most stable structure was already proposed to explain the C chains observed on the Cu(111) surface.¹⁴ However, the most stable chain structure, C-1D(1), has not yet been considered. Both structures are formed entirely by binding sites of type III introduced in Sec. III A.

Interestingly, the other three one-dimensional filaments, C-1D(3), C-1D(4), and C-1D(5), constructed via bonding sites of types I and II, even though they have lower stability, all exhibit the resonance assisted hydrogen bonding (RAHB) effect of Otero *et al.*,¹⁶ where the stabilization energy of the superstructure is greater than the sum of pair binding energies (given by E_{pred}). Notice that the RAHB effect may also be present in the two most stable filaments made of type III binding sites, but this is not clearly seen as $|E_{\text{stab}}| < |E_{\text{pred}}|$ for these structures.

Notice that it is not possible to distinguish the two chains in the STM images by symmetry considerations due to the shape of the molecules. Also note that no chains follow one straight direction for longer than 3–4 C pairs. Since the energies of the two filaments are extremely close, we propose that they both coexist on the surface. We shall assume that the other three filaments are much less likely to occur (their stabilities are considerably smaller), and thus we shall disregard them in the following sections.

TABLE I. Five most stable cytosine filaments formed by periodically repeating two molecules in one dimension. Shown are the two dimers involved [the number in the brackets corresponds to the order of stability according to DFT calculations of dimers (Ref. 43)]; E_{int} , E_{def} , E_{BSSE} , and E_{stab} are the interaction, deformation, BSSE correction, and stabilization energies (including the BSSE corrections), respectively, and E_{pred} is the sum of all dimer energies taken from Ref. 43, and, finally, the length $|\mathbf{A}_1|$ of the lattice vector (in Å) in the direction of periodicity.

Filament	Dimers involved		E_{BSSE}	E_{int}	E_{def}	E_{stab}	E_{pred}	$ \mathbf{A}_1 $
C-1D(1)	$C_5\bar{C}_7(2)$	$C_5\bar{C}_7(2)$	0.39	-2.46	0.57	-1.89	-1.92	9.3
C-1D(2)	$C_7C_7(1)$	$C_5C_5(3)$	0.34	-2.40	0.59	-1.82	-1.86	10.1
C-1D(3)	$C_3\bar{C}_6(6)$	$C_3\bar{C}_6(6)$	0.29	-1.30	0.10	-1.20	-0.86	11.0
C-1D(4)	$C_1C_6(5)$	$C_3C_6(9)$	0.28	-1.22	0.12	-1.09	-0.81	10.6
C-1D(5)	$C_2\bar{C}_6(11)$	$C_2\bar{C}_6(11)$	0.20	-0.61	0.00	-0.61	-0.38	12.1

It is interesting to note that the pairs involved in the two most stable chains (C_7C_7 , C_5C_5 , and $C_7\bar{C}_5$) are also seen in crystalline structures formed by cytosine.⁵⁷⁻⁵⁹ These are the three most stable pairs.⁴³

2. Connecting filaments with each other by end-to-end connections

Although in our calculations we used a periodic model for the filaments, in reality, filaments are of a finite length, and thus can be considered as “larger” molecules (with respect to monomers). Following the same logic as before, these “molecules” can be connected to each other, which results in even bigger structures.

Because of the particular filament structures we find, in which only donor sites of type I are available on either of their long sides, the attachment of two filaments *side-by-side* is impossible by hydrogen bonding. This finding explains why we do not see ordered two-dimensional C islands as the growth of a periodic structure in two dimensions is not possible for this molecule from hydrogen bonding alone (see also Sec. IV). Therefore, only two possibilities remain to attach filaments: by connecting them either (i) end-to-end or (ii) end-to-side, both leading to disordered formations.

To systematically construct all possibilities, we should find all available binding sites for the first two most stable filaments assuming their *finite length*. Both structures C-1D(1) and C-1D(2) contain only type I sites on each *side*. At the ends of the filaments, on the other hand, one can find all types of binding sites from I to III. Since a filament may not contain complete two-molecule cells, four different endings can be found for each of the two filaments. In addition to this, one can flip the filament. To count the total number of possible ends for the two filaments, we note that (i) the filament C-1D(2) is symmetric (its two ends are equivalent) and (ii) when a molecule is removed from either of the ends of C-1D(1), an equivalent end is obtained. Thus, four different endings are available in total before flipping is applied.

By connecting filaments end-to-end, several possibilities exist. We only consider the most stable ones in which binding sites 5 and 7 (see Fig. 3) are involved. In this case, the flipping state of the other filament is fixed by the binding sites used. The eight possibilities are shown in Fig. 5. Each chain can bind to itself in two ways (the other two ways lead

to longer filaments of the same type and thus must be excluded), while two different chains bind to each other in four ways. Some of the chains constructed in this way bend at the connection. The two common trimers which cause the bending in the filaments are also shown, as well as their stabilization energies (in eV) and corresponding sum of dimer energies (in brackets). Note that both energies are rather high, so that these possibilities seem viable.

Since both of the two most stable structures, C-1D(1) and C-1D(2), are possible on the Au(111) surface, we propose that transitions from one structure to the other are also possible within a chain via one of the end-to-end connections considered above. Thus, chains observed in our images may be highly *disordered* as one type of the structure may, at random, be followed by the other.

3. Connecting filaments with each other by T-junction connections

The other possibility is to connect the end of one filament with the side of the other (end-to-side connections), leading to T junctions. Since only type I (two hydrogens) binding sites are available at the sides of either of the two filaments, see Fig. 4, the only possible T junctions result from two chain possibilities (out of four existing filament ends) which have site of type II available (plus flipping possibilities).

If we consider all possible connections to the side of the C-1D(1) filament, eight possibilities arise as site 6 of each chain can bind to either site 2 or 3 along the chain in two chiral ways. Similarly, there are eight possibilities to attach either of the first two most stable filaments to a side of the C-1D(2) structure. All but one of these predicted possibilities are given in Figs. 6(e)–6(k): the one which was not considered as a possibility [namely, C-1D(2) with C-1D(2) via $C_3\bar{C}_6$ bonding] involved overlapping molecules. Note that other possibilities seen in Fig. 6, in particular, (g), (h), (j), and (k), may also be unfavorable due to possible repulsion. Furthermore, these T junctions are only possible to construct if we allow for the oxygen of site 6 to participate in two hydrogen bonds simultaneously, which may considerably reduce the stability of these connections.

In order to understand whether these would be realistic connections, some complementary DFT calculations were performed to check the strength of the hydrogen bonding of

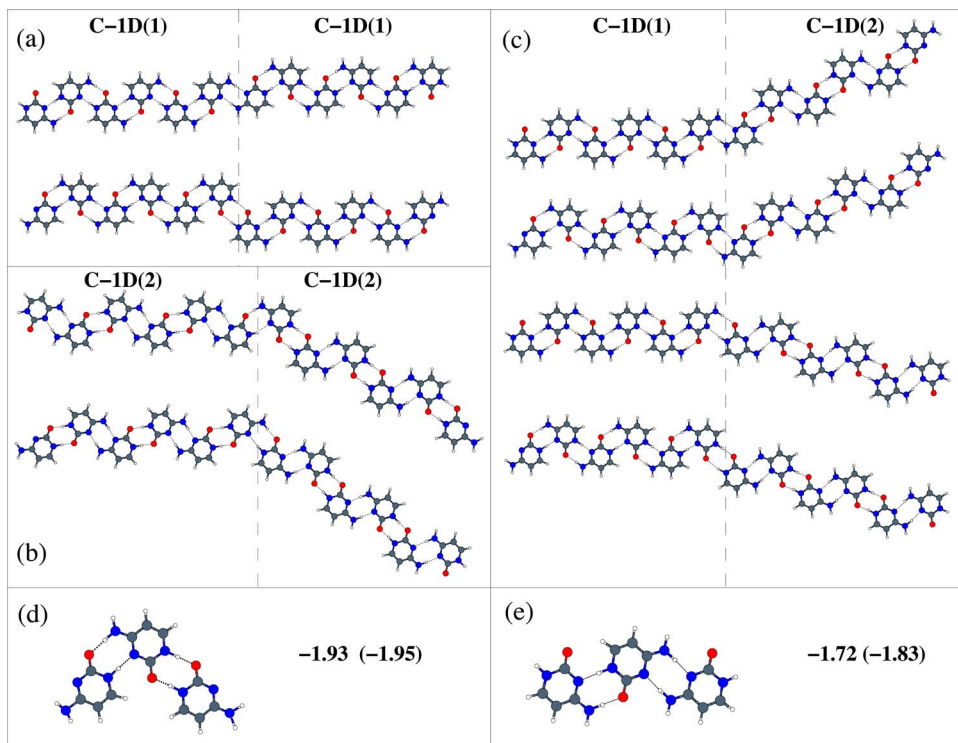


FIG. 5. (Color online) [(a)–(c)] All of the possible end-to-end filament connections. (d) and (e) show the two common trimers which cause the change between (or provide the connection with) two filaments. The calculated stabilization energies of the molecular complexes are also given (in eV), as well as the corresponding sum of dimer energies from Ref. 43 (in brackets).

monomers to chains. All the T-junction possibilities (explained above) were simplified into six possibilities, which are all given in the right part of Fig. 6. In these calculations, we attached a single molecule to the chains in all possible

ways [(a)–(k)]. The stability of the hydrogen bonding which is at work in each particular connection is also given (in eV), as well as the stability of the corresponding isolated pair (in brackets). It is seen that only possibilities involving sites

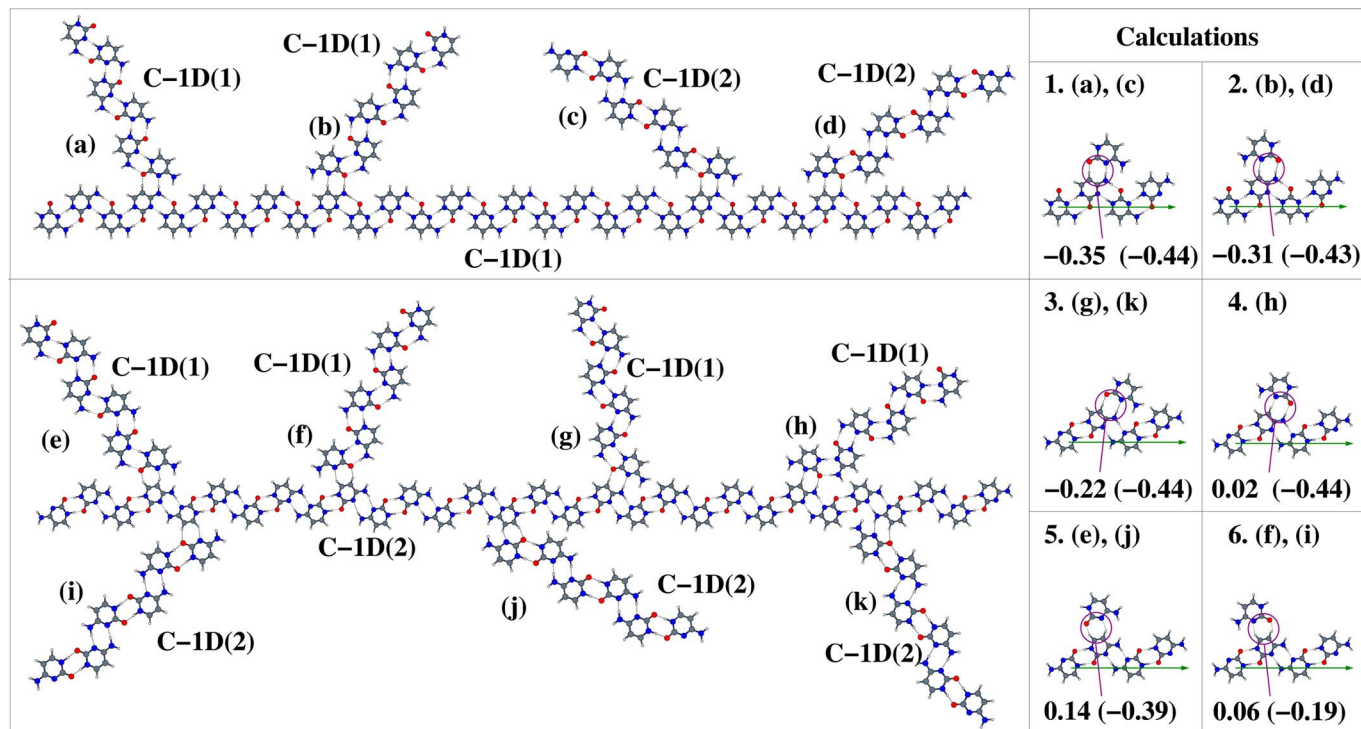


FIG. 6. (Color online) Possible end-to-side filament connections to C-1D(1) [(a)–(d)] and C-1D(2) [(e)–(k)] filaments. Only possibilities associated with the attachment to site 3 of the side of the C-1-D(1) filament are shown in (a)–(d); another four possibilities arise in a similar way when attaching to site 2 of it (not shown). On the right, unit cells of 1D structures (with the lattice vector also indicated, in green) which model the connections employed in all cases [(a)–(k)] are shown; these contain a part of the corresponding filaments (running horizontally) with a single monomer attached. The calculated stabilization energies (in eV) of the relaxed connection structures are also shown together with the stabilization energies of the corresponding individual dimers (in brackets) for comparison. The stabilization energies of the connections were obtained by subtracting the energies of an individual filament and a molecule from the total energy in each case, including the appropriate BSSE correction.

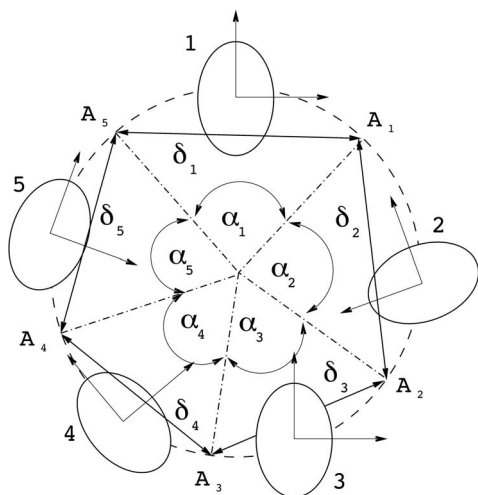


FIG. 7. Schematic of a fivefold ring ($N=5$) with the center at point O (not shown). The segment angles α_i and the corresponding molecular sizes δ_i are explicitly indicated.

3 and 6 (corresponding to the C_3C_6 and $C_3\bar{C}_6$ pairs) are stable and have stabilities comparable to those of the isolated pairs. We anticipate that the calculated stabilization energies of the connections may be slightly modified if more than one molecule was considered connecting to the filaments. However, the simplified models of the connections should be able to give a reasonable estimate of the strength of the T junctions considered. Therefore, the possibilities (e)–(j) and (h) in Fig. 6 cannot be completely discarded, although we believe these are less likely. Note that in cases 4–6, the stabilization energies are positive, which is mainly due to deformation within the monomers; we anticipate then that these particular connections [(h), (f), (e), (i), and (j)] are probably not realized.

4. Ring structures

In order to develop models for five- and sixfold rings observed experimentally (in Fig. 1) we applied a special method similar to that used to construct all possible adenine monolayers.^{27,28} In order to find all rings, we have to consider all possible chains of $N=5$ or 6 molecules in which every molecule forms double hydrogen bonds with both its neighbors along the chain, where the N th molecule is connected to the first one. Using existing information on all C dimers,⁴³ it is possible to construct all possible chains satisfying the above constraints. In practice, we also limited ourselves to only such dimer connections which curve along the chain in the same direction (all are bent either to the right or to the left), facilitating the ring formation. This way, for each possibility, the “key” uniquely identifying the ring is generated as an ordered sequence of all dimers participating in the ring.

Using the suggested method, it was found that, for example, 899 nonequivalent fivefold ring structures are possible. Due to the large number of possibilities generated, it is useful to suggest an approximate way of constructing preliminary ring geometries and propose a simple procedure for estimating their stability—this is what we shall refer to in the

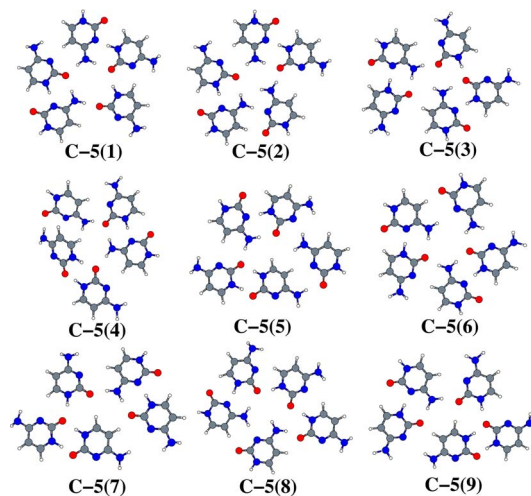


FIG. 8. (Color online) Selected fivefold ring models. These rings are given in order of stability indicated by the number in the brackets.

following as the prediction method; then full DFT relaxation calculations are performed on a selected subset of structures which have the highest (predicted) stabilization energies.

First of all, we shall discuss how the approximate geometries can be generated. Suppose we have N cytosine molecules arranged on a circle of radius R ; each molecule is connected with hydrogen bonds to two of its neighbors on both sides. To arrange the molecules in accordance with their dimer connections, the following procedure was adopted. Each molecule i is given a certain “size” δ_i that corresponds to the distance between two points chosen on either side of the molecule; this parameter depends on the molecule orientation and the binding sites involved. Each point is located 1 Å away from the midpoint between two atoms forming the binding site along the direction perpendicular to the line connecting the two atoms, as shown schematically in Fig. 7. For instance, molecule 1 in the fivefold ring shown in the figure has points A_1 and A_5 on both sides of it, and the distance between them is denoted by δ_1 . Assuming that the ring radius R is large enough, the approximate size of each molecule i can also be written as $\delta_i \approx R\alpha_i$, where α_i is the corresponding segment angle, see Fig. 7. Since all segment angles should sum up to 2π in the circular arrangement, we obtain a simple approximate relationship for the radius,

$$R \approx \frac{1}{2\pi} \sum_{i=1}^N \delta_i. \quad (1)$$

As the value of R is known together with the molecular sizes δ_i , it is not difficult to position the molecules along the circle, which would form an initial geometry of the ring prior to the full-scale DFT calculation.

Also the binding energy of each ring, E_{pred} , can be estimated as the sum of all dimer energies within the ring.

Using these considerations, 899 and 9694 five- and sixfold rings were constructed, respectively. Their predicted energies range between -1.14 and -3.29 eV for the fivefold rings and between -1.34 and -4.29 eV for the sixfold rings.

A total of 41 fivefold ring structures with the highest

TABLE II. Structure and energetics of the cytosine fivefold rings. E_{int} , E_{def} , E_{BSSE} , and E_{stab} are the interaction, deformation, BSSE correction, and stabilization energies (including the BSSE corrections), respectively, and E_{pred} is the sum of all dimer energies taken from Ref. 43. E_{mol} is the stabilization energy per molecule.

Ring	Based on pairs (the ‘‘key’’)					E_{BSSE}	E_{int}	E_{def}	E_{stab}	E_{pred}	E_{mol}
C-5(1)	$C_6\bar{C}_3(6)$	$C_5\bar{C}_7(2)$	$C_6\bar{C}_3(6)$	$C_5\bar{C}_7(2)$	$C_6C_2(7)$	0.70	-4.13	0.80	-3.33	-3.18	-0.66
C-5(2)	$C_6\bar{C}_3(6)$	$C_6C_3(5)$	$C_5\bar{C}_7(2)$	$C_6\bar{C}_3(6)$	$C_5\bar{C}_7(2)$	0.68	-4.03	0.74	-3.29	-3.22	-0.66
C-5(3)	$C_2\bar{C}_6(11)$	$C_1C_6(9)$	$C_7\bar{C}_5(2)$	$C_6\bar{C}_3(6)$	$C_7\bar{C}_5(2)$	0.67	-3.98	0.76	-3.23	-2.67	-0.65
C-5(4)	$C_3\bar{C}_6(6)$	$C_7\bar{C}_5(2)$	$C_3\bar{C}_6(6)$	$C_7C_7(1)$	$C_3\bar{C}_6(6)$	0.68	-4.22	1.02	-3.20	-3.25	-0.64
C-5(5)	$C_6\bar{C}_4(10)$	$C_6\bar{C}_3(6)$	$C_7C_7(1)$	$C_6\bar{C}_3(6)$	$C_5\bar{C}_7(2)$	0.69	-4.26	1.06	-3.20	-3.08	-0.64
C-5(6)	$C_3C_6(5)$	$C_4\bar{C}_6(10)$	$C_7\bar{C}_5(2)$	$C_3\bar{C}_6(6)$	$C_7\bar{C}_5(2)$	0.67	-3.95	0.75	-3.20	-3.06	-0.64
C-5(7)	$C_7C_7(1)$	$C_6C_1(9)$	$C_6\bar{C}_3(6)$	$C_5\bar{C}_7(2)$	$C_6C_3(5)$	0.68	-4.04	0.86	-3.18	-3.19	-0.64
C-5(8)	$C_6\bar{C}_3(6)$	$C_6C_3(5)$	$C_5\bar{C}_7(2)$	$C_6C_1(9)$	$C_6C_1(9)$	0.69	-3.90	0.73	-3.17	-2.57	-0.63
C-5(9)	$C_2\bar{C}_6(11)$	$C_1C_6(9)$	$C_7C_7(1)$	$C_3C_6(5)$	$C_7\bar{C}_5(2)$	0.69	-4.17	1.01	-3.16	-2.95	-0.63

predicted energies were relaxed using the DFT SIESTA method starting from suitable predicted geometries. Nine models are presented in Fig. 8. Interestingly, the rings with C_5 symmetry (not shown) are not the most energetically favorable. The keys of the corresponding structures together with the appropriate energetic information are summarized in Table II. Note that many structures relaxed to the same configuration, the best example being C-5(1), which was the final configuration of four predicted structures. This structure is one of a few possibilities which resemble the shape of a perfect pentagon as observed in experiment (Sec. II). Moreover, some possibilities, for example, C-5(4) and C-5(5), have nonplanar configurations, and therefore were not considered as being suitable candidates for the fivefold rings seen in the experiment.

Although the same method can, in principle, be used for the sixfold rings as well, for these we followed a simpler approach in which each ring was constructed by repeating either a single monomer (C_6 symmetry) or a dimer (C_3 symmetry) along the ring circumference. Since each bond (two bonds) must repeat along the ring, a limited number of possibilities exists, which were then all relaxed using the SIESTA method. Note that each monomer (dimer) is to be rotated to accommodate the rosette formation which results in some stretching of the outer hydrogen bond in every dimer connection. Only three fully relaxed possibilities for the C_6 and two for the C_3 symmetries were found, which are shown in Table III and Fig. 9. Although the ring of C_3 symmetry C-6(2) is the most stable *predicted* structure, it has a triangular shape not observed in our STM images and actually is only the second most stable structure *after* DFT relaxation.

Therefore, we suggest that the most stable *relaxed* ring structure, C-6(1), which has a perfect circular shape, is the model for the experimentally seen sextuplets. Other ring structures in Fig. 9 have much lower stabilization energies.

It is interesting to note that the C-6(1) has exactly the same stability (per molecule) as the fivefold ring C-5(1). Also, for most of the ring structures, the DFT relaxed stabilization energies are larger than the predicted energies implying a small but important role played by the RAHB effect of Otero *et al.*¹⁶ in the C ring structures.

5. Ring junctions

Finally, we discuss the nature of the crossings of branches containing a ring as its vertex. As was discussed in Sec. II, in most cases, three filament branches join to form a sixfold ring. It can easily be seen that a large number of possibilities arises here as well. To illustrate this point, let us consider the most stable circular sixfold ring C-6(1) of C_3 symmetry. The most stable dimers are based on sites 5 and 7 (of type III), and it can be seen that the ring has only three of these sites on its outer perimeter (in fact, all are site 7) which can be used for the branching. Finite chains ending with either sites 5 or 7 can connect to the ring by C_7C_7 or $C_7\bar{C}_5$ bondings. It is clear that a large number of structures are possible, all of very comparable stability. One of these, which may be called a ‘‘roundabout’’ junction, is shown in Fig. 10(d) as an example. Since there are many plausible possibilities which cannot be distinguished in our images (due to the circular shape of the molecules), we propose that all these possibilities are probable. The STM images in Fig. 1

TABLE III. Energies of the cytosine sixfold ring structures. E_{int} , E_{def} , E_{BSSE} , and E_{stab} are the interaction, deformation, BSSE correction, and stabilization energies (including the BSSE corrections), respectively, and E_{pred} is the sum of all dimer energies taken from Ref. 43. E_{mol} is the stabilization energy per molecule.

Ring	Based on pairs		E_{BSSE}	E_{int}	E_{def}	E_{stab}	E_{pred}	E_{mol}
C-6(1)	$C_7\bar{C}_5(2)$	$C_3\bar{C}_6(6)$	0.87	-5.03	0.93	-4.11	-4.17	-0.69
C-6(2)	$C_7C_7(1)$	$C_3C_6(5)$	0.91	-5.46	1.47	-3.98	-4.29	-0.66
C-6(3)	$C_4C_6(8)$	$C_4C_6(8)$	0.66	-4.16	0.62	-3.54	-2.28	-0.59
C-6(4)	$C_6C_1(9)$	$C_6C_1(9)$	0.78	-4.35	0.96	-3.38	-2.23	-0.56
C-6(5)	$C_3C_6(5)$	$C_3C_6(5)$	0.78	-3.61	0.40	-3.21	-2.64	-0.54

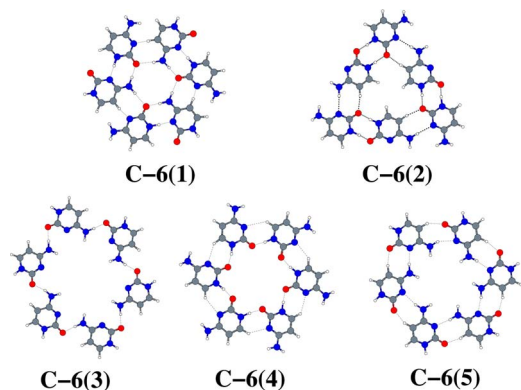


FIG. 9. (Color online) The cytosine sextuplets with three- [structures C-6(1) and C-6(2)] and sixfold [structures from C-6(3) to C-6(5)] symmetries. These rings are given in order of stability indicated by the numbers in the brackets in their label.

show that indeed different geometries exist at the same time on the surface. Some complementary calculations of the rings with one monomer being attached, see Figs. 10(a)–10(c), suggest that many connections may exist, some with considerable stabilization energies. We believe that bonding by site 7 of the rings with sites 7 or 5 at the end of a filament is most plausible.

For the fivefold ring C-5(1), only two *good* sites (5 or 7) are available on its outer perimeter. While three branches may come out of any of the most stable sixfold rings, only two sites of type III exist in any of the most stable fivefold rings (note that some of the rings, e.g., C-5(2) and C-5(5), have three sites 5 and 7 exposed on its perimeter; in all these cases, however, one of the sites 7 has its oxygen atom already involved in the H bond with the neighboring molecule, so that an additional hydrogen bond involving the same O atom would be much weaker than the other two around the perimeter). This finding may explain the experimental observations of Sec. II B that only sixfold rings are capable of serving as a vertex in the C network of branches, while the fivefold rings are met only at the ends of the branches. Very seldom ringlike structures, most of them fivefold rings, which do not have any connection to the branches were detected. We believe these are immobilized by surface imper-

fections. These observations suggest that fivefold rings are stable clusters in their own right, not just branch-terminating structures.

6. Nanocages

Objects inside the nanocages are fast moving species, which makes their identification difficult by STM. However, these must be rather small species to be inside the cages, as seen in the middle of Fig. 4 in Ref. 1, or to diffuse between two cages after a filament has been broken by the tip, offering only a small gate. However, the fact that they should not stick on to any of the surrounding *walls* of the cage (which offer only binding sites of type I) suggests that there must be either none or only weak hydrogen bonds (of type II) available to the species to connect to the walls. Since both most stable filament structures exhibit only donor groups (of type I) on both sides, any cage would also have only these groups available to bind to either from inside or outside of closed cages. A single molecule inside the closed cage structure can still bind to the cage sides via its site 6 (of type II), resulting in the binding of intermediate stability. If a dimer or a more complicated structure (such as a ring, see below) which does not have site 6 available to bind with happens to be trapped inside the cage of a sufficient size during growth, it will not be able to bind to the inside of the cage, and thus will be free to move around the cage to appear in the STM image as a blurred feature. Thus, we believe it is plausible that the blurred features in the nanocages we observe in Fig. 1 are simply small structures unable to bind to the chains.

The five- and sixfold rings and pairs that consume binding sites 5, 6, or 7 (as is actually the case for the most stable pairs) seem to be the most plausible explanation for the blurs, which agrees with the observed ringlike structure found in very narrow nanocages at high coverages, see Fig. 1, and the high mobility of blurs diffusing through narrow gateways between neighboring cages.

IV. KINETICS AND STABILITY

In this paper, we followed a route which is different from the one adopted in studies related to imaging molecular

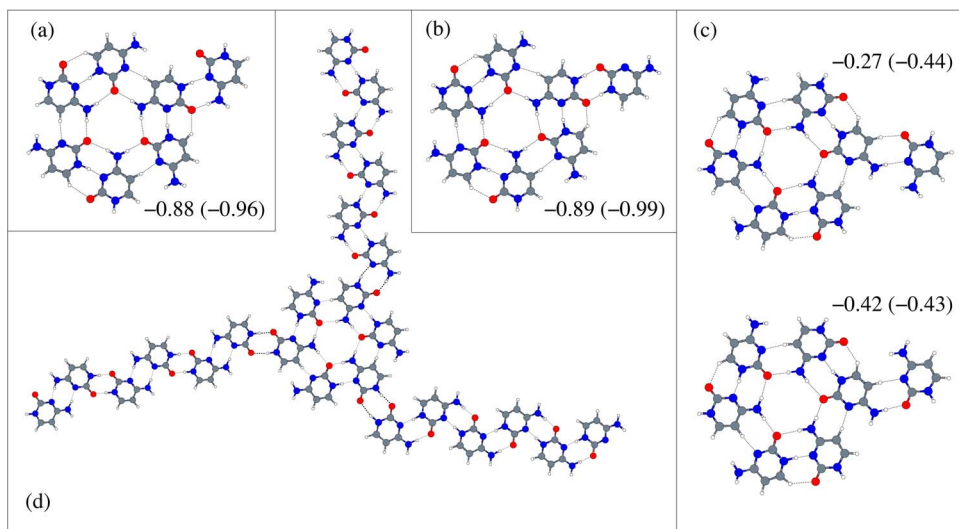


FIG. 10. (Color online) [(a)–(c)] Calculated models for joining molecules to the side of a ring. The stability (in eV) of each pair involved in the ring-monomer binding (which was calculated as the binding energy of a single molecule to the ring, including the BSSE correction) is also given (in eV), together with the stability of each isolated dimer (in brackets) for comparison. (d) A model of the roundabout junction.

assemblies on surfaces with scanning probe methods when one or more theoretical arrangements of molecules which match an experimentally observed image are suggested. Here, this is simply not possible as C forms a disordered structure which may exist in a very large number of geometries. Therefore, our main purpose here was not to suggest an exact match of any of the images, but rather to develop general rules which may be used, if desired, to construct such a match for any part of the experimental image. By using the idea of binding sites and *a posteriori ab initio* calculations, it was possible to show that the observed disorder is a combination of several elementary structures (building blocks or structural motifs): one-dimensional filaments (of which there are 2) and rings (again 2). These interconnect with each other by a final number of junctions, leading to a final variety of the observed geometries. As all the necessary building blocks are known and it is also clear how to connect one with another, we can indeed provide a molecular arrangement which matches any part of the observed STM image, one example of this was given in Ref. 1.

The other question one may ask is why does this particular nucleic acid base molecule form a disordered structure? To address this point, we should consider both kinetic and thermodynamic aspects of the C network formation. In doing so, we shall compare the random network structure discussed above with a 2D periodic (crystallographic) arrangement,⁴ in which 1D periodic filaments lie parallel to each other. In this latter structure, molecules within each filament are connected by hydrogen bonds. However, filaments attach to each other by weak van der Waals (vdW) forces acting between the sides of the filaments. This is similar to the 2D thymine monolayers observed on the Au(111) surface.¹⁷ Note that this ordered C structure has never been observed in our experiments. To compare the stability of the two phases, we should consider their free energies, $F=U-TS$, where U is the internal energy, S entropy, and T temperature. Let us first consider the internal energy U , which will be estimated by the corresponding binding (stabilization) energy. To this end, we consider an island consisting of N parallel filaments each containing N molecules. If E_{Hb} is the energy gain due to hydrogen bonding between two molecules in each filament, and E_{vdW} is the corresponding binding energy due to the vdW interaction between two nearest molecules in adjacent filaments, then the total energy of the island is

$$E_{\text{island}} = N(N-1)(E_{\text{Hb}} + E_{\text{vdW}}) \approx N^2(E_{\text{Hb}} + E_{\text{vdW}}), \quad (2)$$

where in the last passage we assumed $N \gg 1$, which corresponds to neglecting the boundary effects for sufficiently large islands. This energy is to be compared with the average binding energy of a random network containing the same number of molecules, N^2 . In the random network, only at the vertices of the interconnections one molecule is bound to three C molecules; other molecules in filaments and rings have only two neighbors they are bound to via the hydrogen bonds. As a realistic random network is a very complicated system to consider theoretically, we shall adopt a simple model, which, we hope, has all the essential features of the real system. Our model has a single parameter (n , see below) which can be used to balance the number of molecules which

are two- and threefold coordinated. We initially neglect the ring and blur formations and consider a perfect hexagonal lattice containing n molecules on each side of every hexagon excluding the molecules at the vertices. It can easily be seen that the unit cell in this structure consists of $3n+2$ molecules: $3n$ molecules lie on three adjacent filaments plus two molecules at the end vertices. Thus, each unit cell contains two vertex molecules which are threefold coordinated and $3n$ molecules which are twofold coordinated, with the per-cell energy of $E_{\text{cell}} = 3(n-1)E_{\text{Hb}} + 6E_{\text{Hb}} = 3(n+1)E_{\text{Hb}}$. Therefore, the energy of the island containing $M \times M$ repetitions of the unit cell in our model is $E_{\text{network}} = M^2 E_{\text{cell}}$ (neglecting boundary effect, i.e., assuming large M). Since the total number of molecules in the hexagonal island is $M^2(3n+2) = N^2$, we obtain

$$E_{\text{network}} = f_n N^2 E_{\text{Hb}}, \quad (3)$$

where the factor $f_n = (3n+3)/(3n+2) > 1$ for any $n = 0, 1, 2, \dots$. The ratio of the two energies given by Eqs. (2) and (3) is

$$\left| \frac{E_{\text{island}}}{E_{\text{network}}} \right| = \frac{1}{f_n} \left(1 + \frac{E_{\text{vdW}}}{E_{\text{Hb}}} \right). \quad (4)$$

Typically [see, e.g., Fig. 1(c)], $n \sim 6$ (or even larger). Although it is not straightforward to calculate the vdW energy directly, we believe that it can easily be more than 5% of the dimer binding energies which are around $E_{\text{Hb}} \sim -0.9$ eV,⁴³ i.e., $|E_{\text{vdW}}| \geq 0.05|E_{\text{Hb}}|$. This means that $|E_{\text{island}}| > |E_{\text{network}}|$. In other words, at $T=0$ and allowing only filaments to be formed in the random network, we find that the network structure should be less energetically favorable, i.e., we come to the interesting conclusion that the crystalline phase, not the random network, should serve as the *ground state* of the C system at $T=0$.

So far, we have not included blurs and rings into the consideration of the energetics of the network structure. Let us discuss how these structures may affect our conclusion. If we are to assume that blurs contain clusters of C molecules, e.g., rings, then these clusters will mostly contain twofold coordinated molecules. For instance, a ring of m molecules would have the binding energy of mE_{Hb} , which is by E_{Hb} larger than for a filament with the same number of molecules. Therefore, blurs on average increase the number of twofold coordinated molecules and reduce the number of threefold coordinated ones, making our conclusion even stronger. The situation with rings at vertices is not so simple, however. If a ring of m molecules is at the end of a filament, it will contain only one threefold coordinated molecule with the total energy contribution of $(m+1)E_{\text{Hb}}$; if the ring connects two or three filaments, as in Fig. 10(d), two or three molecules will be threefold coordinated with the ring energies of $(m+2)E_{\text{Hb}}$ or $(m+3)E_{\text{Hb}}$, respectively. Therefore, rings with connections to filaments work in the opposite way by increasing the number of threefold coordinated molecules. However, it follows from our images that the number of rings is much smaller than that of the filaments, so that we can expect our conclusion about the prevalence of the crystallographic phase over the random network phase to be still valid.

At nonzero temperatures, the entropy contribution should also be considered. Because there exist a very large number of possible geometries in the random phase, the entropy of this phase should be much larger than that of the crystallographic phase, $S_{\text{network}} \gg S_{\text{island}}$. This means that at high enough temperatures, the random phase should become more stable. Thus, the crystallographic phase is more favorable at low T , while the random phase should be more preferable at large T .

This consideration suggests that during the RT deposition, C molecules will form a random network. This is also easy to understand from the kinetics point of view. Indeed, initially during the deposition, when the coverage is still low, C molecules, freely diffusing on the surface, can easily find other molecules to bind to. Because of the particular arrangement of the binding sites, Fig. 3, where sites 5 and 7, leading to the two most stable dimers, are located on the molecule perimeter in nearly opposite positions, each molecule is preferably connected to two others. Therefore, filaments (up to -0.9 eV per molecule energy gain, Table I) and then rings (up to -0.66 and -0.68 eV per molecule for the five- and sixfold rings, respectively, see Tables II and III) would be formed first. As the deposition continues, and thereby the coverage increases, filaments and rings start to connect to each other, leading to the interconnections described above. As interconnections are formed, the structure at RT stabilizes although it is expected to be still very mobile, especially at low coverages (as mentioned in Sec. II B, at low coverages filaments and blurs may be mobile even at much lower temperatures of 100–150 K at which STM images are taken). Thus, we expect that after deposition the molecules should form a *flexible* random network in which filaments and other small clusters of molecules still remain sufficiently mobile. One may think of the structure formed at RT during deposition as a 2D viscous liquid, at least for not too high coverages.

When the sample is *quickly* taken to the STM held at a rather low temperature of 100–150 K at which molecules are much less mobile, a random network is observed. Although one would expect that at this temperature the system should adopt its ground state, which is the crystallographic phase (this is the case for thymine,¹⁷), apparently this does not happen (assuming, of course, that at these T the entropy contribution should play a much smaller role than at RT). The observed behavior of the C assembly can only be explained by kinetics:¹ when placed in the STM chamber, the random network formed during the deposition does not have enough time to reform into the ground state structure and remains in the structure it accepted at RT. This behavior is similar to the *liquid-glass transition*, so that the observed random network structure may be considered as a glass formed upon rapid cooling of a high temperature 2D liquid phase described above. Note that in the case of thymine,¹⁷ the crystalline arrangement was observed since in this case the vdW interaction between the side groups of the thymine filaments, E_{vdW} , is expected to be much stronger.

One last point to be commented on is related to the number of five- and sixfold rings as a function of coverage (Sec. II B and Fig. 2). At low coverages, dimers would form first,

but there will also be many single molecules as well. Moreover, dimers are expected to be less mobile than single molecules, so that at low coverages stable fivefold rings may form more easily as only five molecules (e.g., two dimers and a single molecule) are required to be in the proximity of each other. However, with the increase in the supply of C molecules at larger coverages and, consequently, reduction of available space, the sixfold rings start to dominate as there will be more dimers available. No more than two branches may connect to the fivefold rings due to the lack of space around the ring and unavailability of appropriate binding sites to match those at the branches ends. At the same time, sixfold rings have a greater variety of available sites around them and thus may easily interconnect with up to three branches. This way, the sixfold rings serve as convenient filaments interconnections and may, in fact, be formed naturally during growth as “meeting points” of several branches. With further increase of coverage, the number of branches cannot increase indefinitely due to lack of space, while the number of interconnections has to increase. Therefore, the number of sixfold rings increases, while the number of fivefold rings does not change any longer.

V. CONCLUSIONS

In this paper, we have discussed the disorder which is formed upon deposition of the nucleic acid base C on the Au(111) surface. This was done using a combination of STM experiments and *ab initio* modeling. In our images, a random network of one-dimensional branches (filaments) and rings which interconnect with each other by various junctions are found. Due to weak interaction of the molecules with the gold surface, the structures formed are stabilized primarily by hydrogen bonding between the molecules. Using *ab initio* DFT calculations, we explain the large number of possibilities observed by the peculiar distribution of binding sites around the molecule which “codes” the disorder during the kinetic process of deposition and growth. The observed disorder occurs due to the development of the basic structures (filaments and rings), which grow from their nucleation points in random directions and join with each other, without the need to reorder to establish a bond.

The process (and thus the structures formed) is entirely random due to almost equal binding energies of various possibilities and is driven by the particular local supply of the C molecules (deposition and surface diffusion) during growth. Some of the smaller structures can be trapped inside nanocages and remain relatively mobile inside them. These smaller structures are most likely C rings or pairs as bigger formations are expected to be less mobile, while single molecules or short branches (trimers, etc.) may still bind to the nanocage walls.

We propose that the ground state of the 2D C assembly is a 2D periodic arrangement in which 1D filaments lie parallel to each other on the surface stabilized by weak vdW interaction between the filaments. However, this phase never realizes upon the fast cooling from the RT random phase as the system gets trapped in the random arrangement and does not have enough kinetic energy to reform during the short

cooling procedure, very much like a liquid-glass transition. It is worth noting that in the presence of a solution, the ordered 2D phase was observed,⁴ which can be explained by greater mobility of the molecules in this case.

Note that we have only included the canonical form of C, yet other tautomeric forms are known to exist.⁶⁰ However, it has been shown that this canonical form produces the most stable configurations for the 3D C crystal structure and some C pairs.⁶⁰ Since C has been seen to form large ordered islands on Au(111),⁴ and also forms ordered 3D crystals,⁵⁷ we believe this effect may be assumed to be minimal.

Finally, as discussed in Ref. 1, such a system is ideal for studying medium-range order in glassy and amorphous materials. Many organic glasses (or amorphous materials) may have underlying elementary structural motifs, which result in the observed long-range disorder when interconnected in a random manner.

ACKNOWLEDGMENTS

We would like to acknowledge the computer time on the HPCx supercomputer provided via the Materials Chemistry Consortium. R.E.A.K. and M.M. are also grateful to the EPSRC for financial support (Grant Nos. GR/P01427/01 and GR/S15808/01). R.O. and M.L. acknowledge support through EU individual Marie Curie Fellowship and through EU-RTN project AMMIST and the EU project PicoInside. We acknowledge the financial support from the Danish Ministry for Science, Technology and Innovation through the iNANO Center, from the Danish Research Councils, from the Danish National Research Foundation, and the Carlsberg Foundation and the EPSRC.

- ¹R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher, *Science* **319**, 312 (2008).
- ²R. E. A. Kelly and L. N. Kantorovich, *J. Mater. Chem.* **16**, 1894 (2006).
- ³W. M. Heckl, D. P. E. Smith, G. Binnig, H. Klagges, T. W. Hansch, and J. Maddocks, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8003 (1991).
- ⁴N. J. Tao, J. A. DeRose, and S. M. Lindsay, *J. Phys. Chem.* **97**, 910 (1993).
- ⁵H. Tanaka, T. Nakagawa, and T. Kawai, *Surf. Sci.* **364**, L575 (1996).
- ⁶T. Wandlowski, D. Lampner, and S. M. Lindsay, *J. Electroanal. Chem.* **404**, 215 (1996).
- ⁷S. J. Sowerby, W. M. Heckl, and G. B. Petersen, *J. Mol. Evol.* **43**, 419 (1996).
- ⁸M. Furukawa, H. Tanaka, and T. Kawai, *Surf. Sci.* **392**, L33 (1997).
- ⁹T. Kawai, H. Tanaka, and T. Nakagawa, *Surf. Sci.* **386**, 124 (1997).
- ¹⁰J. E. Freund, M. Edelwirth, P. Kröbel, and W. M. Heckl, *Phys. Rev. B* **55**, 5394 (1997).
- ¹¹S. J. Sowerby and W. M. Heckl, *Origins Life Evol. Biosphere* **28**, 283 (1998).
- ¹²S. J. Sowerby, M. Edelwirth, and W. M. Heckl, *J. Phys. Chem. B* **102**, 5914 (1998).
- ¹³M. Furukawa, H. Tanaka, and T. Kawai, *Surf. Sci.* **445**, 1 (2000).
- ¹⁴M. Furukawa, H. Tanaka, and T. Kawai, *J. Chem. Phys.* **115**, 3419 (2001).
- ¹⁵Q. Chen, D. J. Frankel, and N. V. Richardson, *Langmuir* **18**, 3219 (2002).
- ¹⁶R. Otero, M. Schöck, L. M. Molina, E. Lægsgaard, I. Stensgaard, B. Hammer, and F. Besenbacher, *Angew. Chem., Int. Ed.* **44**, 2270 (2005).
- ¹⁷W. Xu, R. E. A. Kelly, M. Schöck, R. Otero, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher, *Small* **3**, 2011 (2007).
- ¹⁸L. M. A. Perdigão, P. A. Staniec, N. R. Champness, R. E. A. Kelly, L. N. Kantorovich, and P. H. Beton, *Phys. Rev. B* **73**, 195423 (2006).
- ¹⁹W. Mamdough, M. Dong, R. E. A. Kelly, L. N. Kantorovich, and F. Be-

- senbacher, *J. Phys. Chem. B* **111**, 12048 (2007).
- ²⁰M. Edelwirth, J. Freund, S. J. Sowerby, and W. M. Heckl, *Surf. Sci.* **417**, 201 (1998).
- ²¹T. Nakagawa, H. Tanaka, and T. Kawai, *Surf. Sci.* **370**, L144 (1997).
- ²²J. C. Poler, R. M. Zimmermann, and E. C. Cox, *Langmuir* **11**, 2689 (1995).
- ²³T. Boland and B. D. Ratner, *Langmuir* **10**, 3845 (1994).
- ²⁴K. Ataka and M. Osawa, *J. Electroanal. Chem.* **460**, 188 (1999).
- ²⁵M. Kasaya, H. Tabata, and T. Kawai, *Surf. Sci.* **357-358**, 195 (1996).
- ²⁶K. Shinoda, W. Shinoda, C. C. Liew, S. Tsuzuki, Y. Morikawa, and M. Mikami, *Surf. Sci.* **556**, 109 (2004).
- ²⁷R. E. A. Kelly and L. N. Kantorovich, *Surf. Sci.* **589**, 139 (2005).
- ²⁸R. E. A. Kelly, W. Xu, M. Lukas, R. Otero, M. Mura, J. Y. Lee, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher, *Small* **4**, 1494 (2008).
- ²⁹M. Furukawa, T. Yamada, S. Katano, M. Kawai, H. Ogasawara, and A. Nilsson, *Surf. Sci.* **601**, 5433 (2007).
- ³⁰S. Xu, M. Dong, E. Rauls, R. Otero, T. R. Linderoth, and F. Besenbacher, *Nano Lett.* **6**, 1434 (2006).
- ³¹W. Mamdough, M. Dong, S. Xu, E. Rauls, and F. Besenbacher, *J. Am. Chem. Soc.* **128**, 13305 (2006).
- ³²W. Mamdough, R. E. A. Kelly, M. Dong, L. N. Kantorovich, and F. Besenbacher, *J. Am. Chem. Soc.* **130**, 695 (2008).
- ³³J. V. Barth, *Annu. Rev. Phys. Chem.* **437**, 375 (2007).
- ³⁴J. V. Barth, G. Costantini, and K. Kern, *Nature (London)* **437**, 671 (2005).
- ³⁵S. De Feyter, A. Miura, S. Yao, Z. Chen, F. Wurthner, P. Jonkheijm, A. P. H. J. Schenning, E. W. Meijer, and C. F. De Schryver, *Nano Lett.* **5**, 77 (2005).
- ³⁶S. De Feyter and C. F. De Schryver, *Chem. Soc. Rev.* **32**, 139 (2003).
- ³⁷D. L. Keeling, N. S. Oxtoby, C. Wilson, M. J. Humphry, N. R. Champness, and P. H. Beton, *Nano Lett.* **3**, 9 (2003).
- ³⁸J. A. Theobald, N. S. Oxtoby, M. A. Phillips, N. R. Champness, and P. H. Beton, *Nature (London)* **424**, 1029 (2003).
- ³⁹T. Kawasaki, K. Suzuki, Y. Hakoda, and K. Soai, *Angew. Chem.* **47**, 496 (2008).
- ⁴⁰N. J. Tao and Z. Shi, *J. Phys. Chem.* **98**, 1464 (1994).
- ⁴¹M. Komiyama, M. Gu, T. Shimaguchi, H.-M. Wu, and T. Okada, *Appl. Phys. A: Mater. Sci. Process.* **66**, S635 (1998).
- ⁴²D. J. Frankel, Q. Chen, and N. V. Richardson, *J. Chem. Phys.* **124**, 204704 (2006).
- ⁴³R. E. A. Kelly, Y. J. Lee, and L. N. Kantorovich, *J. Phys. Chem. B* **109**, 22045 (2005).
- ⁴⁴P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik, and J. Vondrasek, *J. Comput. Chem.* **18**, 1136 (1997).
- ⁴⁵R. E. A. Kelly, Y. J. Lee, and L. N. Kantorovich, *J. Phys. Chem. B* **109**, 11933 (2005).
- ⁴⁶R. E. A. Kelly and L. N. Kantorovich, *J. Phys. Chem. B* **110**, 2249 (2006).
- ⁴⁷L. Petersen, M. Schunack, B. Schaefer, T. R. Linderoth, P. B. Rasmussen, P. T. Sprunger, E. Lægsgaard, I. Stensgaard, and F. Besenbacher, *Rev. Sci. Instrum.* **72**, 1438 (2001).
- ⁴⁸E. Lægsgaard, O. Osterlund, P. Thosttrup, P. B. Rasmussen, I. Stensgaard, and F. Besenbacher, *Rev. Sci. Instrum.* **72**, 3537 (2001).
- ⁴⁹J. V. Barth, H. Brune, G. Ertl, and R. J. Behm, *Phys. Rev. B* **42**, 9307 (1990).
- ⁵⁰S. Rapino and F. Zerbetto, *Langmuir* **21**, 2512 (2005).
- ⁵¹P. Ordejon, E. Artacho, and J. M. Soler, *Phys. Rev. B* **53**, R10441 (1996).
- ⁵²J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, *J. Phys.: Condens. Matter* **14**, 2745 (2002).
- ⁵³J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ⁵⁴R. E. A. Kelly and L. N. Kantorovich, *J. Phys. Chem. C* **111**, 3883 (2007).
- ⁵⁵S. F. Boys and F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- ⁵⁶S. Piana and A. Bilic, *J. Phys. Chem. B* **110**, 23467 (2006).
- ⁵⁷D. L. Barker and R. E. Marsh, *Acta Crystallogr.* **17**, 1581 (1964).
- ⁵⁸M. Mirzaei, F. Elmi, and N. L. Hadipour, *J. Phys. Chem. B* **110**, 10991 (2006).
- ⁵⁹J. P. Garcia-Teran, O. Castillo, A. Luque, U. Garcia-Couceiro, G. Beobide, and P. Roman, *Cryst. Growth Des.* **7**, 2594 (2007).
- ⁶⁰Z. Yang and M. T. Rodgers, *Phys. Chem. Chem. Phys.* **6**, 2749 (2004).

The Journal of Chemical Physics is copyrighted by the American Institute of Physics (AIP). Redistribution of journal material is subject to the AIP online journal license and/or AIP copyright. For more information, see <http://ojps.aip.org/jcpo/jcpcr/jsp>