Surface-Induced Keto–Enol Tautomerization of DNA Base Molecules and Consequent [4 + 2]-like Cycloaddition on Si(111)7×7

Lei Zhang, Hanieh Farkhondeh, Fatemeh Rahnemaye Rahsepar, Avisek Chatterjee, and Kam Tong Leung*

ABSTRACT: Adsorption and film growth of deoxyribonucleic acid (DNA) base molecules (cytosine, guanine, thymine, and adenine) on Si(111)7×7 have been studied by combining X-ray photoelectron spectroscopy (XPS) with ab initio calculations based on the density functional theory (DFT). Multiple tautomeric forms and keto–enol tautomerization are revealed by the O 1s, N 1s, and C 1s XPS spectra of the O-containing DNA bases: cytosine, guanine, and thymine. While the carbonyl group-containing keto tautomer is more stable in a thick film and in powder, the hydroxyl group-containing enol tautomer is found at the interface. The keto–enol tautomerization, as induced by the reactive Si(111)7×7 surface, leads to the formation of a conjugated aromatic six-membered ring with a delocalized π electron system and to the consequent [4 + 2]-like cycloaddition between the enol tautomer and the 7×7 surface. The DFT calculation suggests that the enol tautomer exhibits a kinetic advantage over the keto one for the [4 + 2]-like cycloaddition. Among the several plausible pathways for the cycloaddition provided by the enol tautomer, the experimentally determined one involves a ring N and ring C atom (a polar pair), rather than two ring C atoms (a nonpolar pair), to better match the polar Si adatom–restatom pair of the 7×7 surface. Furthermore, the reacted ring C atom does not have any attached terminal functional group (e.g., −NH2 and −OH). Further deposition leads to continuous film growth in the keto tautomeric form for cytosine and guanine. For the only O-free DNA base molecule, adenine, active bonding N → Si, rather than the [4 + 2]-like cycloaddition, is observed on the 7×7 surface. Of the four DNA base molecules, adenine is also the only one with its aromaticity maintained when adsorbed on the Si(111)7×7 surface. A reactive surface like the 7×7 surface could therefore provide a new control to trigger tautomerization that is often associated with genetic mutation.

INTRODUCTION

Functionalization of semiconductor surfaces with biomolecules, such as deoxyribonucleic acid (DNA), proteins, and peptides, has attracted much recent attention because of their potential applications in biosensors, biochips, and bioelectronics.1–4 These organic–inorganic hybrid systems can be used to incorporate versatile biochemical properties of biomaterials into well-developed inorganic microelectronic components, thereby creating novel devices. The chemical, electronic, and structural properties of these hybrid systems greatly depend on the chemical bonding and molecular assembly of the biomolecules on the semiconductor surfaces. However, full characterization of the macro-biomolecular semiconductor interface is often challenging because of the complexity of the macro-biomolecules and the considerable number of possible reaction pathways and surface processes that could occur on the semiconductor surface. As DNA base molecules and amino acids are the simpler constituents of DNAs and proteins, better understanding of the adsorption of these building-block biomolecules on semiconductor surfaces could provide important insights into more complex macro-biomolecular processes. To date, most of the published results regarding adsorption of DNA bases were acquired on single-crystal noble metal surfaces, such as Au(111),5–10 Cu(110),11–20 and Cu(111),21–23 or on a Au foil24 under ultrahigh vacuum condition or on graphite25,26 under ambient or solution environment. There are only a limited number of similar experiments that involved semiconductor substrates.27–31 For DNA bases on single-crystal metal surfaces, where biomolecule-to-metal atom interactions are found to play a minor role, self-assemblies of one-dimensional chains or two-dimensional arrays due to the stronger interactions among biomolecules have often been observed by scanning tunneling microscopy (STM). Calculations based on density functional theory (DFT) further show that in most cases dimers, rather than individual molecules, are the building blocks of these self-assembled DNA base structures. In a dimer, two DNA base molecules are doubly or triply hydrogen bonded to each other.32 Since each DNA base molecule contains both...
Reconstructed semiconductor surfaces, characterized by localized directional bonding bands, tend to form covalent bonding with organic molecules. Several types of interfacial reactions have been well established for organic molecules on two of the most studied reconstructed Si surfaces, Si(100)2×1 and Si(111)7×7. They include proton-transfer reactions (e.g., N-H and O-H dissociation), [2 + 2]-like and [4 + 2]-like cycloaddition, ene reaction, and dative bond reaction.3,34 The experimentally observed reaction products are the results of combined influences of both thermodynamic and kinetic factors. With cytosine and thymine being pyrimidines and guanine and adenine being purines, the four DNA base molecules all contain aromatic ring structures with multiple functional groups. The aforementioned types of reactions are, in principle, all possible for these DNA base molecules upon functional groups. The surface-induced keto–enol tautomerization leads to the formation of a conjugated aromatic six-membered ring, which consequently facilitates the [4 + 2]-like cycloaddition between the enol tautomer and the 7×7 surface. This conjugated aromaticity is kinetically favorable for the cycloaddition, as suggested by our DFT calculation of the [4 + 2]-like cycloaddition of cytosine in both enol and keto forms through the same pathway. On the other hand, the O-free base (adenine) is found to adsorb on the 7×7 surface by dative bonding between a pyridinic N atom and a Si adatom. Of the four DNA bases, adenine is therefore the only one with its aromaticity maintained at the interface.

Experimental Details. The experiments were performed in a multi-chamber ultrahigh vacuum system (Omicron Nanotechnology), with a base pressure lower than 5 × 10⁻11 mbar, described elsewhere.35 The analysis chamber was equipped with an XPS spectrometer consisting of a SPHERA hemispherical electron analyzer with a 7-Channeltron detector assembly and a monochromatized Al Kα source (1486.7 eV photon energy), along with a variable-temperature scanning probe microscope for atomic-resolution imaging. Single-side polished Si(111) chips (11 × 2 mm², 0.3 mm thick) with a resistivity of 5 mΩ cm (dopant: arsenic, orient: (111) ± 0.9°, Virginia Semiconductors) were used as the substrates. Large terraces of the 7×7 surface reconstruction were observed by STM, after outgassing the Si substrate at 400 °C overnight followed by flash-anneling at 1200 °C several times by direct-current heating. An example of such an STM image of 7×7 surface is shown in Figure S1 (Supporting Information). Deposition of the biomolecules was performed in an adjoining ultrahigh vacuum preparation chamber. Powders of cytosine, thymine, adenine (all 99% purity), and guanine (98% purity, all from Sigma-Aldrich) were separately loaded into individual effusion cells (designed for low-temperature evaporation of organic materials, Dr. Eberl MBE-Komponenten GmbH) and outgassed thoroughly before exposure. The effusion cell temperatures used for the deposition of cytosine, guanine, thymine, and adenine (with their respective normal melting points of 325, 360, 317, and 365 °C) on the 7×7 substrate were 150, 240, 108, and 115 °C, respectively. The molecular identities of these DNA base molecules during deposition were confirmed by their cracking patterns, collected in situ with a quadrupole mass spectrometer and found to be in good accordance with the available references.38 Unless stated otherwise, XPS spectra of the Si 2p, O 1s, N 1s, and C 1s regions were collected with a pass energy of 20 eV, at which the peak width for the Ag 3d₅/₂ photoline at 368.3 eV achievable with our spectrometer was 0.7 eV full width at half-maximum (FWHM). All spectra were fitted with mixed Gaussian–Lorentzian line shapes (70% Gaussian and 30% Lorentzian) along with a Shirley background using Casa-XPS software. The binding energies are referenced to the Si 2p₃/₂ photoline at 99.3 eV of bulk Si. We have also measured the core-level spectra of the raw powders of the DNA bases, in which case an electron neutralizer was employed to compensate for the minor charging during the XPS measure-
The binding energy scale of the powder spectra was calibrated with respect to that of the corresponding thick films by aligning the respective main N 1s feature.

The first-principle DFT calculations were performed using the plane-wave based Vienna Ab initio Simulation Package (VASP, Version 5.4) implemented on the Materials Exploration and Design Analysis (MedA, version 2.19, Materials Design, Inc.) platform. The projector-augmented wave method\textsuperscript{39,50} was used to describe the electron–ion interactions, and the generalized gradient approximation with the exchange–correlation functional as defined by Perdew–Burke–Ernzerhof\textsuperscript{51} was used to model the electron–electron exchange–correlation energy. The dispersion energies were included in the calculations, as implemented within the DFT-D3 formalism.\textsuperscript{52,53} Conjugate gradient algorithm was employed to optimize the geometry of the atomic structure, and all Si atoms were completely relaxed until the forces on all the atoms were less than 0.05 eV/Å. The plane-wave expansion cutoff energy was set to 400 eV, and the surface Brillouin zone was sampled at the Γ point with a k-point spacing of 0.5 Å\textsuperscript{−1}. The energy convergence of the self-consistent field was set to \(1.0 \times 10^{-5}\) eV, with a Methfessel–Paxton smearing of 0.2 eV. The dimer-adatom-stacking fault model was used to model the Si(111)\textsuperscript{7} surface.\textsuperscript{54} A periodic repeating slab consisting of two Si bilayers and a reconstructed adatom–restatom topmost layer (with a total number of 200 Si atoms) with a lattice constant of 5.41 Å and a vacuum gap of 10 Å was used to represent the Si(111)\textsuperscript{7} surface, and the bottom layer of the Si slab was terminated by 49 H atoms. To find the most stable adsorption configurations of the O-containing base molecules (cytosine, guanine, and thymine) following the \([4 + 2]\)-like cycloaddition, we placed the adsorbates for each available pathway on a center adatom–restatom pair inside a faulted half unit cell. For adenine, we placed the adsorbate on a center Si adatom in a faulted half unit cell with N – Si dative bonding for each pyridinic N. The positions of all atoms and molecules were relaxed during the DFT-D3 calculations. The adsorption energy, \(E_{\text{ad}}\), is obtained with the formula \(E_{\text{ad}} = E_{\text{M-slab}} - E_{\text{slab}} - E_{\text{M}}\), where \(E_{\text{M-slab}}\), \(E_{\text{slab}}\), and \(E_{\text{M}}\) are the total energies of a base molecule adsorbed on the Si\textsubscript{200}H\textsubscript{48} slab, the Si\textsubscript{200}H\textsubscript{48} slab (as the model \textsuperscript{7}x\textsuperscript{7} surface), and the isolated base molecule, respectively.

**RESULTS AND DISCUSSION**

Figure 1 shows the O 1s, N 1s, and C 1s spectra of cytosine films deposited on Si(111)\textsuperscript{7}x\textsuperscript{7} with the effusion cell held at 150 °C for a total of 180, 360, 720, and 3840 s (corresponding to a nominal coverage of 1.5 \(\times\) 10\textsuperscript{−2}, 3.0 \(\times\) 10\textsuperscript{−2}, 6.0 \(\times\) 10\textsuperscript{−2}, and 3.2 \(\times\) 10\textsuperscript{−1}\) Langmuir, respectively) as well as those of the powder sample. The general resemblance of the core-level spectra for the powder and the as-grown 3840 s film suggests that cytosine remains in the keto tautomer form in the thick film because in the solid phase (powder), cytosine is known to exist in the major tautomeric form (i.e., the keto form).\textsuperscript{39} Ball-and-stick models of the keto and enol tautomers of an isolated cytosine molecule obtained by DFT calculations.

![Figure 1. XPS spectra of the O 1s, N 1s, and C 1s regions of cytosine powder and of cytosine films deposited on Si(111)7x7 with the effusion cell held at 150 °C for 180, 360, 720, and 3840 s, along with ball-and-stick models of the keto and enol tautomers of an isolated cytosine molecule obtained by DFT calculations.](https://doi.org/10.1021/acs.langmuir.1c03173)
electrons localized only at the two double bonds. In other words, the enol tautomer displays a character of aromaticity, while the keto one does not. Previous studies of chemisorption of conjugated aromatic molecules on the $7 \times 7$ surface have shown that the most favored adsorption reaction is the $[4 + 2]$-like cycloaddition, in which two weak $\pi$ bonds are replaced by two strong $\sigma$ bonds.\textsuperscript{35−38} When compared with the $[2 + 2]$-like cycloaddition, the energy cost due to the strain of the bonded molecules (caused by dimension mismatch between the molecule and a Si−Si pair on the surface) for the $[4 + 2]$-like cycloaddition has been found to be smaller. For the cytosine enol tautomer, the $[4 + 2]$-like cycloaddition with a Si−Si pair could occur through pathways involving three different sets of $\pi$ bonds: (1) N1=C2 and N3=C4; (2) N3=C4 and C5=C6; and (3) C5=C6 and N1=C2. As both N−Si bonding and pyridinic-N N 1s features are found to have a similar binding energy near 399.1 eV,\textsuperscript{38,57−60} the N 1s spectrum is not helpful in determining whether the pyridinic N is involved in the $[4 + 2]$-like cycloaddition. With smaller intrinsic widths found for individual C 1s components than those for the O 1s and N 1s features, the C 1s spectrum can be used to deduce the reaction pathway by examining the binding energy shifts of relevant C features, resulting from bond breakage and formation. A cursory comparison of the C 1s spectra of the 3840 s film and 180 s film shows the following differences: (a) the two valleys between C6 and C5 and between C4 and C6 in the spectrum of the 3840 s film have disappeared in that of the 180 s film, and (b) the edge at the high binding energy side of the entire C 1s band for the 3840 s film has shifted to the lower binding energy side in that for the

<table>
<thead>
<tr>
<th>DNA base</th>
<th>O 1s</th>
<th>N 1s</th>
<th>C 1s</th>
<th>binding energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cytosine</td>
<td>O−H</td>
<td>C≡O</td>
<td>−NH−</td>
<td>−NH\textsubscript{2}</td>
</tr>
<tr>
<td>thick film</td>
<td>S31.7</td>
<td>400.7</td>
<td>399.9</td>
<td>399.1</td>
</tr>
<tr>
<td>thin film</td>
<td>S32.9</td>
<td>400.7</td>
<td>399.9</td>
<td>399.1</td>
</tr>
<tr>
<td>guanine</td>
<td>O−H</td>
<td>C≡O</td>
<td>−NH−</td>
<td>−NH\textsubscript{2}</td>
</tr>
<tr>
<td>thick film</td>
<td>S31.6</td>
<td>400.6</td>
<td>399.7</td>
<td>399.0</td>
</tr>
<tr>
<td>thin film</td>
<td>S32.9</td>
<td>400.7</td>
<td>399.7</td>
<td>399.0</td>
</tr>
<tr>
<td>thymine</td>
<td>O−H</td>
<td>C≡O</td>
<td>−NH−</td>
<td>==N−</td>
</tr>
<tr>
<td>thick film</td>
<td>S32.2</td>
<td>400.9</td>
<td>399.7</td>
<td>399.0</td>
</tr>
<tr>
<td>thin film</td>
<td>S33.0</td>
<td>531.5</td>
<td>400.7</td>
<td>399.0</td>
</tr>
<tr>
<td>adenine</td>
<td>−NH−</td>
<td>−NH\textsubscript{2}</td>
<td>==N−</td>
<td>C6</td>
</tr>
</tbody>
</table>

Figure 2. Equilibrium geometries and adsorption energies following $[4 + 2]$-like cycloaddition of a cytosine molecule in its (a) keto or (b–d) enol form to a center Si adatom−restatom pair in the faulted half unit cell of the Si(111)$7 \times 7$ surface through different pathways: (a) keto via pathway (2): N3=C4 and C5=C6; (b) enol via pathway (2): N3=C4 and C5=C6; (c) enol via pathway (1): N1=C2 and N3=C4; and (d) enol via pathway (3): C5=C6 and N1=C2. AA-Si and RA-Si correspond to the Si center adatom and Si restatom of the $7 \times 7$ surface, respectively.
180 s film, while the edge at the low binding energy side does not show any shift. For the C 1s spectrum of the 180 s film, we have fitted the spectrum with four features (with the same widths) at 285.0, 286.0, 287.5, and 288.5 eV, which are attributed to C5, C6, C4, and C2, respectively. Except for the C5 feature, the binding energies of these C 1s components all shift by 0.5–0.7 eV to a lower binding energy relative to those for the thick film. The unchanged binding energy for C5 therefore rules out the [4 + 2]-like cycloaddition through the C5=C6 and N1=C2 pathway, which otherwise would lead to the formation of C5–Si and C2–Si σ bonding and accordingly a decrease in the binding energy of C5 by ~0.8 eV.61,63

Furthermore, the lower binding energy for C2 by 0.5 eV is not related to C2–Si bonding but to the keto–enol tautomerization, in which −NH−C2−O is converted to −N==C2−OH. The binding energy decrease in both C6 and C4 could only be accounted for if the [4 + 2]-like cycloaddition occurs through the N3=C4 and C5=C6 set, in which the binding energy of C6 is lowered by 0.7 eV due to conversion of the C5=C6 bond to a C5–C6 and a C6–Si bonds, while the binding energy of C4 is lowered by 0.5 eV due to conversion of the C5−C4=N3 moiety to the C5−C4=N3 moiety (in which one C–N bond is in effect replaced by one C–C bond). Further deposition of cytosine (to 360, 720, and 3840 s) leads to intensity increase in the keto tautomer-related features in both the spectra of O 1s (at 531.7 eV for carbonyl O) and N 1s (at 400.7 eV for pyrrolic N), suggesting that the keto tautomer of cytosine grows gradually into a thick film on top of the interfacial enol-dominant adlayer.

It is of interest to understand why the enol tautomer is preferred over the keto tautomer for the [4 + 2]-like cycloaddition at all, when the same cycloaddition via pathway (2) could also proceed in the keto form. We therefore calculate the adsorption energies of both cytosine tautomers after the [4 + 2]-like cycloaddition through the same N3=C4 and C5=C6 pathway [pathway (2)]. We also determine the adsorption energies of the [4 + 2]-like cycloaddition of the enol tautomer through the other two pathways (the [4 + 2]-like cycloaddition through these two pathways cannot take place for the keto tautomer). The adsorption energies and equilibrium geometries of the most stable adsorption configurations are shown in Figure 2. Surprisingly, the adsorption energy of the keto tautomer (−154.59 kJ/mol) obtained by the [4 + 2]-like cycloaddition through the pathway (2) is found to be more negative than that of the enol tautomer (−105.69 kJ/mol). This is not in accordance with our experimental observation provided by XPS results, which therefore suggests that the observed preference of enol to keto tautomer in the [4 + 2]-like cycloaddition is driven kinetically rather than thermodynamically. The unique delocalized π electron system in the enol tautomer could provide some plausible advantages kinetically (e.g., smaller energy barriers, low-energy intermediates) in initiating and/or facilitating the [4 + 2]-like cycloaddition. On the other hand, among the three pathways available for the enol tautomer, the experimentally detected one [that is, through the pathway (2)] has considerably lower adsorption energies than the other two (−105.69 kJ/mol vs −58.52 and −59.83 kJ/mol) and is therefore thermodynamically more favorable.

Figure 3 shows the O 1s, N 1s, and C 1s spectra of guanine films deposited on Si(111)7×7 with the effusion cell held at 240 °C for 240, 480, 960, and 3840 s (corresponding to a nominal coverage of 3.6 × 10−2, 7.2 × 10−2, 1.4 × 10−1, and 5.8 × 10−1 Langmuir, respectively) as well as those of guanine powder. The O 1s and N 1s spectra of the thick 3840 s film and the powder are consistent with the guanine keto tautomer, being the most stable tautomer in the solid phase. The single O 1s feature at 531.6 eV corresponds to the carbonyl O, while the three N 1s features at 399.0, 399.7, and 400.6 eV with an intensity ratio of 2:1:2 can be assigned to pyridinic, amino, and pyrrolic N, respectively. The three well-resolved C 1s features observed at 285.3, 286.5, and 288.1 eV with an intensity ratio of 1:2:2 can be attributed to C5, (C4, C8), and (C2, C6), respectively, based on the consideration of electronegativities of their nearest neighbors. The differences between the binding energies for C4 and C8 and between those for C2 and C6 are too small to be discerned. Like cytosine, guanine also undergoes chemisorption on the 7×7 surface in the form of enol tautomers instead of keto tautomers. This is supported by the predominant hydroxyl O feature at 532.9 eV in the O 1s spectrum of the 240 s film. Furthermore, the enol tautomer could come directly from the evaporated guanine molecules, which consist of both keto (60%) and enol (40%) tautomers,43 and/or from surface-induced keto–enol tautomerization of the keto tautomer upon adsorption.

The [4 + 2]-like cycloaddition is expected to occur between the guanine enol tautomer and a Si–Si pair (of the 7×7 surface) through one of four pathways involving different sets of double-bond pairs: (1) N1=C6 and C5=C4; (2) C5=C4 and N3=C2; (3) N3=C2 and N1=C6; and (4) C4=C5 and N7=C8. For the 240 s film, the broad C 1s spectrum can be fitted with five components with the same FWHM at 285.1, 286.1, 286.6, 287.4, and 288.0 eV. The C 1s component at 285.1 eV could readily be assigned to C5 with the same binding energy as that of the thick film, suggesting an unreacted C5 site and thus ruling out pathway (2). The C 1s component at 287.4 eV is attributed to C6 with a 0.6 eV lower in binding energy than that of the thick film due to keto–enol tautomerization, as similarly found for that in cytosine and thus ruling out pathway (3). The C 1s component at 288.0 eV can be assigned to C2 with the same binding energy as that of the thick film, suggesting an unreacted C2 site and thus again ruling out pathway (2). The two remaining components at 286.1 and 286.6 eV correspond to C4 and C8, both of which have a binding energy of 286.5 eV for the thick 3840 s film. This suggests that one of these C sites has a binding energy 0.4
C4 and C5 and a new C4-C5 double bond is replaced by one C4-C5 single bond and a new C4-Si bond upon guanine adsorption. C6 and C2 are not viable sites for cycloaddition, likely because of unfavorable effects caused by the respective presence of terminal hydroxyl and amino functional groups attached to the ring at these C sites. For the 240 s film, the main N 1s feature at 399.0 eV corresponds to pyridinic N3 and N7 and to N1-Si bonding. The intensity of this feature is about three times that of the pyrrolic N9-H feature at 400.7 eV and that of amino N at 399.7 eV. Further deposition leads to a continuous growth of guanine adlayers in the keto form, as supported by further intensity increases in the carbonyl O and pyrrolic N features attributed to the keto tautomer, which is similarly found for the adsorption and film growth of cytosine on the 7×7 surface.

Figure 4 shows the O 1s, N 1s, and C 1s spectra of thymine deposited on Si(111)7×7 with the effusion cell held at 108 °C for 10, 60, and 600 s (corresponding to a nominal coverage of 9.8 × 10⁻², 5.9 × 10⁻¹, and 5.9 Langmuir, respectively), along with those of thymine powder. The similarities in the O 1s and N 1s spectra between the thick 600 s film and the powder are consistent with the di-keto tautomer of thymine as the most stable form in the solid phase. The predominant O 1s feature at 532.2 eV and N 1s feature at 400.9 eV therefore correspond to the carbonyl O and pyrrolic N in the di-keto tautomer, respectively. The four C 1s features at 285.4, 286.5, 288.6, and 289.7 eV are readily attributed to (C5, C7), C6, C4, and C2, respectively, based on the electronegativities of their nearest neighbors.

We have also calculated the adsorption configurations and energies of the [4 + 2]-like cycloaddition through the four aforementioned pathways between an enol guanine and a Si adatom–restatom pair, as displayed in Figure 4. Evidently, pathways (1) and (3) involving N and C atoms (Figure 4a,c) lead to more stable adsorption than pathways (2) and (4) involving two C atoms (Figure 4b,d), probably due to polarity matching between N–C and Si–Si pairs. On the other hand, the experimentally determined cycloaddition pathway (1) through N1 and C4 is less thermodynamically stable than the pathway (3) through N3 and C6 (with an adsorption energy of −59.38 kJ/mol for the former and −90.70 kJ/mol for the latter), suggesting that pathway (1) is kinetically more favorable. An apparent difference between C4 [involved in pathway (1)] and C6 [involved in pathway (3)] in an unreacted enol guanine molecule is that all bonds to C4 are located in the flat pyrimidine–imidazole ring, while one bond of the hydroxyl O to C6 is outside of the flat pyrimidine ring, which might prevent C6 in pathway (3) from approaching a Si restatom due to the steric hindrance. Another difference between the equilibrium geometries of the reacted guanine molecule following pathways (1) and (3) is that in the former case, the pyrimidine–imidazole ring is located totally inside a potential well centered at the bonded restatom and surrounded by three Si adatoms, while in the latter case, the imidazole ring is outside of the potential well. The structural matching of the former could provide a kinetic advantage for the initiation and proceeding of the cycloaddition reaction. In addition, pathway (4) involving the imidazole ring (Figure 4d) has a positive adsorption energy, and is therefore not feasible, likely due to the larger strain caused by the dimension mismatch upon cycloaddition.

Figure 5 shows the O 1s, N 1s, and C 1s spectra of thymine deposited on Si(111)7×7 with the effusion cell held at 108 °C for 10, 60, and 600 s (corresponding to a nominal coverage of 9.8 × 10⁻², 5.9 × 10⁻¹, and 5.9 Langmuir, respectively), along with those of thymine powder. The similarities in the O 1s and N 1s spectra between the thick 600 s film and the powder are consistent with the di-keto tautomer of thymine as the most stable form in the solid phase. The predominant O 1s feature at 532.2 eV and N 1s feature at 400.9 eV therefore correspond to the carbonyl O and pyrrolic N in the di-keto tautomer, respectively. The four C 1s features at 285.4, 286.5, 288.6, and 289.7 eV are readily attributed to (C5, C7), C6, C4, and C2, respectively, based on the electronegativities of their nearest neighbors. The binding energy of the carbonyl O 1s in the
thymine di-keto tautomer is found, remarkably, to be 0.5 eV higher than those in the cytosine and guanine keto tautomers. This could be due to the presence of two carbonyl groups in the thymine di-keto tautomer, each of which shares less valence electrons in the pyrimidine ring than the single carbonyl case in cytosine and guanine, therefore resulting in a higher binding energy.

The predominant hydroxyl O 1s feature at 533.0 eV in the 10 s film, together with the minor pyrrolic N 1s feature at 400.7 eV, shows that thymine occurs as the di-enol tautomer form at the interface. In contrast to cytosine and guanine, which consist of both keto and enol tautomers in the gas phase, the di-keto tautomer has been reported to be the only stable tautomer in the gas phase of thymine.45,64 The 7×7 surface therefore facilitates surface-induced di-keto-to-di-enol tautomerization, followed by [4 + 2]-like cycloaddition that occurs between the di-enol tautomer and the 7×7 surface. For a di-enol tautomer with a conjugated aromatic ring, the [4 + 2]-like cycloaddition could take place through one of three pathways involving (1) N1=C2 and N3=C4, (2) N3=C4 and C5=C6, and (3) C5=C6 and N1=C2. The C 1s spectrum of the 10 s film could be fitted with four features at 285.4, 286.2, 288.0, and 288.6 eV, which are assigned to (C5, C7), C6, C4, and C2, respectively. Similar to the cytosine adsorption on the 7×7 surface, the binding energies of all the C 1s species, except for the (C5, C7) feature, become smaller in the interfacial adlayer than those in the thick (600 s) di-keto tautomer.

For the 60 s film, the carbonyl O 1s feature at 531.5 eV and the pyrrolic N 1s feature at 400.7 eV become more intense. The binding energy of the carbonyl O 1s feature at 531.5 eV is, however, 0.7 eV smaller than that of the di-keto tautomer but is similar to those of cytosine and guanine keto tautomers. The adsorbed thymine in the 60 s film therefore occurs in the mono-keto-mono-enol tautomer form, in which tautomerization takes place only for one of the two carbonyl-pyrrolic groups (Figure 5). In this mono-keto-mono-enol tautomer, the H=N1−C2=O moiety has changed to N1==C2−O−H because the C2 feature for the di-keto tautomer at 289.7 eV remains absent in the C 1s spectrum of the 60 s film. Since the O 1s and N 1s features for the mono-keto-mono-enol tautomer are negligible for the 10 s film and they only become evident for the 60 s film, the mono-keto-mono-enol tautomer could adsorb on the interfacial di-enol tautomers (that are present upon cycloaddition), forming a second adlayer that acts essentially as a transitional adlayer between the interfacial adlayer and multilayers of the thick film. This transitional adlayer is found to remain intact after annealing at 100 °C for 600 s, while the di-keto multilayers cannot even survive an overnight storage at room temperature (not shown). The thermal stability of the transitional adlayer is believed to result from hydrogen bonding between the transitional and interfacial adlayers and is consistent with similar thermal stability found for other amino acids on Si(111)7×7.65 It is, however, unclear why this mono-keto-mono-enol tautomer has not been adopted in the transitional adlayer rather than the di-keto tautomer. We hypothesize that structural matching between the hydrogen bond donors and acceptors could play a crucial role. For the 60 s film, the intensity of the hydroxyl O 1s feature is about three times that of the carbonyl O 1s feature. This intensity ratio found for the two O 1s features is therefore consistent with the coverage of the interfacial adlayer in the di-enol tautomer form (each with two hydroxyl O) being similar to that of the transitional adlayer in the mono-keto-mono-enol tautomeric form (each with one carbonyl O and one hydroxyl O).

Our above analysis therefore shows that all three O-containing DNA base molecules (cytosine, guanine, and thymine) react with Si(111)7×7 surface via the [4 + 2]-like cycloaddition, similar to that found for the adsorption of other aromatic molecules with six-membered rings, including benzene, pyridine, pyrimidine, and triazine, on Si(111)-7×7.35−38 The reaction pathways of these DNA base molecules exhibit three common features. (1) Only the enol tautomer is involved in the cycloaddition. The enol tautomer is already present partly in the gas phase (cytosine and guanine) and/or is induced by the reactive 7×7 surface (cytosine, guanine, and thymine). Unlike the keto tautomer, the enol tautomer is a conjugated ring system with alternating single and double bonds in the ring and delocalized π orbitals, which could facilitate the cycloaddition. (2) The bonding to the Si atoms (of the 7×7 surface) is via one N and one C atom rather than two C atoms. This could be related to the polarity of the six-membered ring with N being electron-rich and C being electron-deficient, which matches electrostatically with the electron-deficient adatom electron-rich restatom pair on the 7×7 surface. An electrostatic interaction could therefore play an important role kinetically to initiate the [4 + 2]-like cycloaddition. (3) Among the few candidates for such N and C pairs, the preferred pair is characterized by the absence of any attached functional group (such as −OH or −NH₂).
Furthermore, the N atoms used to form the N–Si bonding, that is, N3 for cytosine and thymine and N1 for guanine, indeed correspond to those used to form H bonding in the canonical DNA base pairing. The C atoms used to form the C–Si bonding, that is, C6 for cytosine and thymine and C4 for guanine, are all involved in C==C double bonds rather than C==O or C==N double bonds. It is noteworthy that tautomerization and the associated formation of minor tautomers in the DNA base molecules are responsible for the occurrence of non-canonical base pairing, for example, guanine–adenine pair. This rare base mismatch is the origin of spontaneous mutation because the integrity of genomic information is not maintained. A distribution of multiple tautomers for a DNA base molecule (i.e., tautomeric equilibrium) has been reported to be affected by physical and chemical controlling factors such as pH, temperature, presence of metals, or UV light exposure. The present result suggests a reactive surface with directional dangling bonds (such as the 7×7 surface) could offer a new control of inducing tautomerization of O-containing DNA base molecules.

Figure 6 depicts the N 1s and C 1s spectra of adenine films deposited on Si(111)7×7 with the effusion cell held at 115 °C for 60, 120, 480, and 1920 s along with a ball-and-stick model of an isolated adenine molecule obtained by DFT calculation.

![Figure 6](https://doi.org/10.1021/acs.langmuir.1c03173)

**Figure 6.** XPS spectra of the N 1s and C 1s regions of adenine powder and of adenine films deposited on Si(111)7×7 with the effusion cell held at 115 °C for 60, 120, 480, and 1920 s, along with a ball-and-stick model of an isolated adenine molecule obtained by DFT calculation.

for 60, 120, 480, and 1920 s (corresponding to a nominal coverage of 5.0 × 10⁻², 9.9 × 10⁻², 4.0 × 10⁻², and 1.6 Langmuir, respectively) along with those of adenine powder. For the 1920 s film and adenine powder, the three N 1s features at 399.0, 399.6, and 400.6 eV correspond to pyridinic, amino, and pyrrolic N, respectively, and the observed intensity ratio of 3:1:1 is in good accordance with their stoichiometric ratio. The binding energies of these N 1s species of this O-free DNA base molecule are consistent with those of the other three (O-containing) DNA base molecules. The C 1s spectrum for the 1920 s film can be fitted with five components with the same FWHM (1.0 eV) at 285.3, 286.2, 286.5, 286.7, and 287.3 eV, which are attributed to C5, C4, C2, C8, and C6, respectively, in accordance with the presence of one N or two N nearest neighbors and with reference to the binding energies of C 1s in guanine (with a similar structure, Figure 3). The C 1s spectrum for the adenine powder is similar to that of the thick film, except for the discernibly larger intensity found for the component at 285.3 eV, likely due to the contribution of minor contaminants in the powder caused by ambient handling.

In the N 1s spectrum of the 60 s film, an additional N 1s feature at 401.1 eV is required to provide a good fit to the spectrum. The sum of the intensities of this additional feature and pyridinic N 1s feature is three times the intensity of the amino or pyrrolic N 1s feature. This suggests that some pyridinic N is converted into this new moiety. The observed 2.1 eV increase in the N 1s binding energy from that for pyridinic N is consistent with the dative bonded pyridinic N. The donation of its lone-pair electrons in a pyridinic N to an electrophilic Si adatom could give rise to a considerable increase in the N 1s binding energy. As an essentially Lewis acid–base interaction, this N → Si dative bonding has also been reported to account for pyridine adsorption on Si(111)7×7 and trimethylamine adsorption on Si(100)-2×1.

As the lone-pair electrons of pyridinic N are located on the ring plane, a dative bonding with the Si adatom would suggest that the adenine ring plane has become vertical to the Si surface. It is not possible, however, to determine which one of the three pyridinic N atoms (N1, N3, and N7) could form the dative bonding from the N 1s and C 1s spectra alone. Our DFT calculation, as shown in Figure S3 in Supporting Information, confirms that the planar pyrimidine–imidazole ring is vertical to the 7×7 surface, and that the dative bonding via N3 has the lowest adsorption energy. While the [4 + 2]-like cycloaddition occurred in the three O-containing DNA base molecules breaks the aromaticity of the base molecules, dative bonding in adenine retains the aromaticity. Reactivity index and orbital hardness have been computed for the DNA base molecules using DFT calculations to evaluate the reactivity of each atom and the stability of aromaticity of each DNA base molecule. The presence of oxygen atom has been found to decrease the stability of aromaticity of the base molecules, while the O-free adenine has the most stable aromaticity among the four DNA base molecules, which is in agreement with our XPS results.

### CONCLUSIONS

The formation of different tautomers and keto–enol tautomerization have been found to play important roles in the adsorption and film growth of O-containing DNA base molecules (cytosine, guanine, and thymine) on the Si(111)7×7 surface. Indeed, while the keto tautomer is the only form in thick films and powders, the presence of the enol tautomer at the interface is especially significant in silicon surface chemistry. Furthermore, keto–enol tautomerization induced by the Si(111)7×7 surface is found unambiguously for thymine and most likely also for cytosine and guanine. The formation of the enol tautomer, with its conjugated six-membered ring, appears to be a prerequisite for the [4 + 2]-like cycloaddition between the DNA base molecule and the 7×7 surface. From our DFT calculation, the enol tautomer is preferable to the keto one in terms of kinetics of the [4 + 2]-like cycloaddition. The resulting cycloaddition adsorption product involves two strong σ bonds formed between N and C ring atom sites (of the base molecule) without any functional group attached and the Si atoms of the 7×7 surface. Further deposition leads to continuous film growth in the keto tautomer form for cytosine and guanine, but for thymine, a transitional adlayer in the mono-keto-mono-enol tautomer form occurs before the growth of the di-keto multilayers. For the O-free DNA base, adenine, dative bonding of pyridinic N
to Si, rather than the \([4 + 2]\)-like cycloaddition, appears to be the preferred route. Of all four DNA base molecules, adenine is the only one whose aromaticity is preserved when adsorbed on the Si(111)7\times 7 surface, in agreement with its most stable aromaticity predicted by an earlier theoretical study.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.langmuir.1c03173.

STM image of the flashed-annealed Si(111)7\times 7 surface used as the substrate, and equilibrium geometries and adsorption energies for a thymine molecule at a center Si adatom–restatom pair site and for an adenine molecule at a center Si adatom site, all obtained by DFT calculations (PDF).

## AUTHOR INFORMATION

### Corresponding Author

Kam Tong Leung — WATLab, and Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; orcid.org/0000-0002-1879-2806; Email: tong@uwaterloo.ca

### Authors

Lei Zhang — WATLab, and Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Hanieh Farkhondeh — WATLab, and Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; orcid.org/0000-0001-6670-3452

Fatemeh Rahnamaye Rahsepar — WATLab, and Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; School of Chemistry, College of Science, University of Tehran, Tehran 1417614411, Iran; orcid.org/0000-0003-2359-2523

Avisek Chatterjee — WATLab, and Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.langmuir.1c03173

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The work was supported by the Natural Sciences and Engineering Research Council of Canada.

## REFERENCES


(69) Cao, X.; Hamers, R. J. Silicon surfaces as electron acceptors: Dative bonding of amines with Si(001) and Si(111) surfaces. J. Am. Chem. Soc. 2001, 123, 10988−10996.