Self-assembly of Natural and Unnatural Nucleobases at Surfaces and Interfaces

Artur Ciesielski, Mohamed El Garah, Stefano Masiero, Paolo Samorì

WILEY-VCH

DOI: 10.1002/ ((please add manuscript number))
Article type: Review

Self-assembly of natural and unnatural nucleobases at surfaces and interfaces

*Artur Ciesielski, Mohamed El Garah, Stefano Masiero and Paolo Samorì**

Dr. A. Ciesielski, Dr. M. El Garah, Prof. P. Samorì
ISIS & icFRC, Université de Strasbourg & CNRS
8 allée Gaspard Monge, 67000 Strasbourg (France)
E-mail: samori@unistra.fr
Homepage: http://www.nanochemistry.fr

Prof. S. Masiero
Dipartimento di Chimica "G. Caimician"
Alma Mater Studiorum – Università di Bologna
v. San Giacomo, 11 – 40126 Bologna (Italy)

Keywords: self-assembly, nucleobases, scanning tunnelling microscopy, hydrogen bonding

Abstract

The self-assembly of small organic molecules interacting *via* non-covalent forces is a viable approach towards the construction of highly ordered nanostructured materials. Among various molecular components, natural and unnatural nucleobases can undergo non-covalent self-association to form supramolecular architectures with *ad hoc* structural motifs. Such structures, when decorated with appropriate electrically/optically active units, can be used as scaffolds to locate such units in pre-determined positions in 2D on a surface, thereby paving the way towards a wide range of applications, e.g. in opto-electronics. This *Review* discusses some of the basic concepts of the supramolecular engineering of natural and unnatural nucleobases and derivatives thereof as well as the self-assembly processes on conductive solid substrates, as investigated by scanning tunnelling microscopy in ultra-high vacuum and at the solid/liquid interface. By unravelling the structure and dynamics of these self-assembled architectures with a sub-nm resolution a greater control over the formation of more and more sophisticated functional systems is achieved. The ability to understand and predict how nucleobases interact both among themselves, as well as with other molecules is extremely important, since it provides access to ever more complex DNA- and RNA-based nanostructures and nanomaterials as key components in nanomechanical devices.

1. Introduction

Attaining an exquisite control over the position and organization of molecules into monolayers on solid surfaces with a nanoscale precision represents a major step towards the fabrication of multifunctional nanodevices. The self-assembly of small organic molecules interacting via non-covalent forces is a practical method for developing highly ordered nanostructured materials. Among non-covalent interactions, hydrogen bonding offers great control over the process of molecular self-assembly[1] because it combines selectivity, directionality, reversibility and cooperativity. Such a unique character is the basis of sophisticated programs for self-assembly such as those relying on the Watson–Crick base pairing[2] that govern the generation of complex architectures like the fascinating DNA double helix. The formation of duplex DNA/RNA from their single stranded components is the result of panoply of non-covalent intermolecular interactions, such as π-stacking, van der Waals forces, and hydrophobic effects. Nevertheless, the high fidelity observed in the binding patterns of complementary DNA/RNA sequences is mainly due to the unique molecular recognition capability of naturally occurring nucleic acid bases (nucleobases, NB) through hydrogen-bonding interactions. The five natural nucleobases, i.e. adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) (**Figure 1a**), are involved in the self-assembly of one of Nature's most fascinating class of biopolymers, i.e. DNA and RNA. The A and G components are purine derivatives, while C, T and U nucleobases are pyrimidine derivatives. The use of the common NB in supramolecular chemistry offers the flexibility of exploiting four different binding units A, C, G and T (or U), all of which offer different binding characteristics. The two major NB binding motifs present in nucleic acids, i.e. guanine-cytosine (G•C), and adenine-thymine (A•T) (adenine-uracil, A•U in RNA) are portrayed in **Figure 1b**. These NB pairs interact *via* 3 and 2 H-bonds, respectively. While Watson-Crick base pairing is governing the association of nucleic acids, the self-assembly of NBs is not limited to Watson-Crick type of interactions. In fact, there are nearly 30 possible patterns of

base-pairs, that involve at least two H-bonds, which can be formed between the five common NBs.[3] While excellent reviews have been published on the molecular recognition *via* base pairing and the use of NBs as supramolecular motifs,[4] recent progress in supramolecular engineering of the NB-based 2D self-assembled architectures on various conductive solid substrates prompted us to author this *Review* article.

Scanning tunnelling microscopy (STM)[5] became a widely explored tool to investigate molecular structures and assemblies at interfaces with a resolution of nm, offering direct insight into the 2D world of non-covalent interactions.[6] When adsorbed on graphite, the STM current images show a brighter contrast for conjugated moieties and darker for aliphatic groups. Such a contrast is ruled by the resonant tunnelling between the Fermi level of the substrate (in this case graphite) and the frontier orbital of the molecules. Roughly speaking, the energy difference between them is inversely proportional to the tunnelling probability.[7] The spatial sub-nanometer resolution that can be achieved by STM imaging allows one to gain detailed insight into molecular interactions. STM is therefore the tool of choice to assist the design of molecular modules that can undergo programed self-assembly at surfaces under precise conditions. Presently, the STM exploration of molecular adsorption can be studied under various environmental conditions, e.g. solid/liquid interface and ultra-high vacuum (UHV). The former offers an exceptionally attractive environment for the self-assembly of small organic building blocks, and has several advantages, if compared to experiments performed under UHV: (*i*) the experimental approach is straightforward and does not demand complex infrastructures; (*ii*) the dynamic adsorption-desorption process fosters self-healing of defects present the layers physisorbed at the solid-liquid interface;[8] (*iii*) it offers an ideal environment for *in situ* chemical modifications of adsorbed species. STM operating at the solid/liquid interface offers the possibility of screening the changes in the structural motif of molecular monolayers when external physical or chemical stimuli are applied, e.g. varying the pH[9] or by coordination of metallic centers to organic species.[10] Such external modification

can occur in the reversible manner, while under UHV molecular re-organizations is mostly irreversible.

In this *Review* article, we will discuss the engineering of supramolecular structures formed through self-association of natural and unnatural nucleobases on atomically flat solid substrates as explored by STM operating at the solid/liquid interfaces and under UHV. The first section (2.1) will be devoted to the recent progress in self-assembly of natural NBs, which primary use mono- and hetero-NBs as either H-bonding motif or as metal-binding ligands. In the second section (2.2) we will present and discuss systems relying on the formation of H-bonds between unnatural NBs.

2. Self-assembly in 2D

By exploiting the *bottom-up* approach, supramolecular chemistry provides a rational path to the design of molecules capable of undergoing self-assembly at surfaces forming pre-programmed structures. On the nanometer scale, molecules are the favourite building blocks to decorate, structure, and functionalize surfaces. Self-assembly towards the formation of a targeted 2D structure, is governed by the subtle balance between molecule–substrate, molecule–molecule, solvent–substrate and molecule–solvent interactions. While negligible under UHV conditions, the effect of the chosen solvent on the supramolecular assembly at the solid/liquid interface has been explored by numerous researchers over the past years.[11] The organic solvents employed for STM measurements need to combine certain characteristics: (*i*) being electrochemically inert under experimental conditions, (*ii*) possess a low vapour pressure enabling measurements to be carried out with the tip immersed inside one drop of solution (ca. 5–20 μL), without the necessity of employing a sealed fluid cell, (*iii*) be a good solvent for the compound under study, and (*iv*) have a low affinity for the substrate, i.e. have a low tendency to adsorb on its surface**.** Atomically flat surfaces are ideal substrate to have a high control over the molecular self-assembly process as they make it possible to steer the

molecular ordering *via* specific geometric and electronic effects. The typical conductive substrate employed in STM measurements is highly oriented pyrolitic graphite (HOPG), which can be freshly produced by cleavage using scotch tape. Metallic surfaces like Ag(111), Au(111), Cu(110) and Cu(111) are the substrate of choice for measurements performed in UHV.

At the solid/liquid interface, interactions between molecules, solvent and the substrate are essential in defining the supramolecular architectures. These interactions constitute one aspect of a complex thermodynamic description of the self-assembly process, which necessarily also includes parameters such as temperature, entropy, or chemical potentials. Tremendous efforts have been made to find appropriate thermodynamic and kinetic models that quantitatively describe experimental results, and have been reviewed recently by Gutzler and Rosei.[12] Several studies show that kinetically stabilized phases can form at a surface, which over time transform into thermodynamically more stable polymorphs. Nevertheless, so far, a conclusive and universal thermodynamic description is still missing. Furthermore, as in any reaction, thermodynamic and kinetic factors influence the formation of supramolecular structures. Thermodynamics of molecular self-assembly, and in particular the enthalpy and the entropy, can be modulated by varying experimental conditions. By maximizing the concentration of the solution, the adsorbate density is maximized, thus the gain in enthalpy is maximized. However, molecular physisorption at surface is accompanied by a loss in (translational) entropy, occurring through the minimization of degrees of freedom of the system. The influence of entropy can be minimized *via* proper design of molecules, e.g. by incorporating rigid units to decrease the number of available molecular conformations.[13] The scenario becomes more complicate when dealing with a solution featuring two components to be co-deposited at the surface, to form mixed polymers/architectures. If one of the two components is already physisorbed on the surface, its partial desorption may be energetically unfavored, thereby hindering the emergence of molecular recognition leading to homogeneous

intermixing on the substrate surface. Because of this reason upon use of a high concentrated solution frequently only one component adsorbs at the solid/liquid interface. To have two components co-deposited at surface it is mandatory to use solutions with an overall molecular concentration lower than the one needed to form a monolayer.

2.1. Two-dimensional self-assembly of natural nucleobases

2.1.1 Guanine

Among nucleobases, guanine (G) is the most interesting and versatile one, because the presence of N(1)-H and N(2)-H donor- and O(6), N(3) and N(7) acceptor-sites located in a self-complementary arrangement coupled with a polarized aromatic surface allows it to self-assemble through H-bonding to give a variety of supramolecular architectures.[14] Depending on the experimental conditions it can undergo different self-assembly pathways. In the presence of certain metal ions, guanines can form G-quartet (hereafter G_4) based architectures (**Figure 2**) such as octamers or columnar polymeric aggregates, stabilized by cyclic N(2)-H···N(7) and N(1)-H···O(6) H-bonds and coordination bonds between guanine and metal ion. In the absence of metal templating centers, guanines can self-assemble, both in solution and in the solid state, into ribbon-like architectures (Figure 2).

Two different ribbons, exhibiting different H-bond patterns can be formed in the solid state and in solution, i.e. G-ribbon A (Figure 2), which is thermodynamically stable in the solid state and can be detected in solution in (anhydrous) chloroform soon after dissolving the polycrystalline powder, is characterized by cyclic N(2)-H···O(6) and N(1)-H···N(7) H-bonds. In solution the G-ribbon A slowly undergoes a structural transition towards a thermodynamically more stable G-ribbon B (Figure 2), characterized by N(1)-H···O(6) and N(2)-H···N(3) intermolecular cyclic H-bonds. Upon adsorption on solid substrate, guanine supramolecular ribbons undergo a back rearrangement into the A-type ribbons. This adaptive supramolecular behaviour makes guanine the most studied nucleobase.

Self-assemblies generated via hydrogen bonds

Since the pioneering work on the STM visualization of guanines on MoS_2 and HOPG by Heckl and co-workers in 1991,[15] the self-assembly behaviour of G molecules has been extensively studied by various research groups. In 2005 Besenbacher and co-workers demonstrated,[16] by high-resolution variable-temperature STM, that guanine (***G1***) (Figure 3) deposited under ultraclean conditions onto an inert Au(111) substrate self-assembles into H-bonded $\boldsymbol{G1}_4$-based 2D networks. After a few years,[17] it has been shown that upon deposition of ***G1*** molecules onto Au(111) at room temperature (RT), a heterochiral phase corresponding to two enantiomerically pure homochiral $\boldsymbol{G1}_4$-based 2D networks (R and L) is formed (**Figure 4a**). Within the $\boldsymbol{G1}_4$-based architecture, the molecules interact *via* N(2)-H···N(7) and N(1)-H···O(6) H-bonding and form $\boldsymbol{G1}_4$. Moreover, because of the presence of N(9) hydrogen, the molecules interact *via* N(9)-H···N(7) H-bonds, and ultimately form $\boldsymbol{G1}_4$-based 2D networks. Interestingly, upon annealing at 400 K a new heterochiral intermixed $\boldsymbol{G1}_4$-based architecture is formed consisting of equal amounts of ***G1*** molecules in the chiral R and L form. Noteworthy, different H-bonding patterns of various NBs have been listed in **Table 1**.

Methylation of NB is an important control mechanism in biology which is applied, for example, in the regulation of gene expression.[18] The effect of methylation on the intermolecular interactions between G molecules was recently studied by Wang and co-workers.[19] The STM analysis, corroborated by density functional theory (DFT), revealed that methylation of guanine can have subtle effects on both pairing nature and the strength of H-bonds, with a strong dependence on the position of methylation. In particular, the formation of H-bonded ribbons at the solid/liquid interface can be achieved by methylation of G in N(9) position (***G2***).

Recently, functionalization of N(9) position of G molecules was employed by Xu and co-workers[20] to steer the self-assembly of G into ribbon-like architectures under UHV

conditions. While it was considered that G molecules self-assemble into G_4-based 2D networks on Au(111), the authors demonstrated that N(9)-ethylguanine (***G3***) once deposited on gold surface forms ***G3***-ribbons A (Figure 4b). Noteworthy, upon co-deposition of ***G3*** and Fe ions onto Au(111), isolated $\boldsymbol{G3}_4$ were observed (Figure 4c).

To achieve an *in-depth* understanding of the self-assembly of guanine at the solid/liquid interface, we performed a sub-molecularly resolved STM study of physisorbed monolayers on graphite of a series of N(9)-alkylated guanines with linear alkyl side-chains from $–C_2H_5$ up to $–C_{18}H_{37}$ (***G3***-***G11***).[21] This comparative study was carried out by applying a drop of a solution of the chosen alkyl substituted guanine in 1,2,4-trichlorobenzene (TCB) on freshly cleaved HOPG surface. The presence of a long aliphatic side chain was expected to promote both solubility of G molecules in the organic solvent and the molecular physisorption on HOPG. All G derivatives were found to form monomorphic 2D crystals, which were stable on the several minutes timescale. Minor changes of the alkyl side-chains length drastically impact on the 2D pattern morphology (**Figure 5**). The derivatives with alkyl tails consisting of at least 12 carbon atoms (***G8***-***G11***) were found to self-assemble into linear H-bonded ribbons through the N(2)-H···O(6) and N(1)-H···N(7) pairing, with each unit cell containing four molecules (Figure 5f-5i). An identical H-bonding motif was observed for N(9)-ethylguanine ***G3*** (Figure 5a), but the packing shows only two molecules per unit cell. No H-bonded supramolecular polymers were monitored in the case of ***G4***-***G7*** (tails from C_6 to C_{10}), at the surface: ordered monolayers of single rows of (non-H-bonded) molecules (Figure 5b and 5e) or H-bonded dimers (Figure 5c and 5d) were rather observed, in which the formation of ordered self-assembled structures is driven by both intramolecular and molecule-HOPG van der Waals interactions.

The self-assembly of H-bonded networks of a lipophilic G derivative can be utilized to design highly ordered supramolecular structures.[22] For instance, micrometer-long molecule-thick ribbons can be grown on a mica surface from deposits of lipophilic G derivatives

bearing two alkyl groups (***G13***). This arises from the self-assembly into highly directional A-type ribbons, ultimately forming 2D polycrystalline structures of parallel ribbons at the solution/graphite interface.[23] This latter structure reflects the supramolecular motif that has been detected both in the single crystal[24] and as a metastable state in solution by NMR spectroscopy.[23] Moreover this architecture is of interest for its ability to rectify currents, making it a potential building block for the construction of nanoscale bio-electronic devices and circuits.[25]

In an effort to adjust and improve the electronic properties of guanosine derivatives, we extended our scope to investigate C(8)-substituted lipophilic oxoguanosine derivative ***G14***.[26] The cooperative effect of H-bonding and solvophobic interactions induces the C(8)-oxoguanosines to self-assemble into helical architectures both in the liquid crystalline phase, in solution and at the solid/liquid interface. These arrangements, which are markedly different from the one generated through the self-association of G derivatives unsubstituted in the C(8) position, are of interest for their unique optical properties.

Guanine tautomerisation at surface

Tautomerisation, i.e. a ubiquitous phenomenon characterized by transfer of a hydrogen atom or proton, has been found to extensively exist in N-heterocyclic compounds such as G.[27] In biological systems, the transformation of nucleobases from canonical to their non-canonical forms could induce mismatch of base pairing and further disturb the genetic codes.[28] Direct real-space evidence on existence of different G tautomers and further investigation on the effect of metals on G tautomerisation at surfaces has been recently reported by Xu and co-workers.[29] From the interplay of STM imaging under UHV and DFT calculations, the authors show that tautomerisation of G from G/N(9)-H (***G1***) to G/N(7)-H (***G12***) is facilitated on Au(111) surface by annealing, whereas such tautomerisation process is

effectively inhibited by introducing Ni atoms due to its preferential coordination at the N(7) site of G/N(9)-H tautomer.

Assembly/reassembly processes on surfaces

Lately, we have reported on the sub-nanometer scale K^+-templated reversible assembly/reassembly process of ***G11*** into highly ordered ***G11***$_4$'s and ***G11***-ribbons A (**Figure 6**).[30] The formation of ***G11*** supramolecular structures has been studied on graphite, was followed by addition of potassium picrate ($K^+(pic)^-$), cryptand[2.2.2], and trifluoromethanesulfonic acid (HTf) in order to trigger the reversible interconversion between two different highly-ordered supramolecular architectures, i.e. ribbon and G_4. The monolayer of a neat ***G11*** displays a crystalline pattern composed of ***G11***-ribbons A (Figure 6). Upon *in situ* addition of 10 mM potassium picrate solution to the ***G11***-ribbons in Figure 6a, the ***G11***$_4$-based architecture was generated (Figure 6b). The addition of a [2.2.2]cryptand solution to the ***G11***$_4$ pattern on graphite, resulted in the reassembly of ***G11*** into ***G11***-ribbons (Figure 6c). By varying the pH with HTf, as a result of the K^+ release from the cryptate, and the ***G11***$_4$ assembly was restored (Figure 6d). Upon subsequent addition of a solution containing an excess of [2.2.2]cryptand, the ***G11***-ribbons were regenerated (Figure 6e). This demonstrates the potential of G-based structures to behave as a 2D dynamer,[31] whose response to external chemical stimuli can be monitored by STM on the sub-nanometer scale in real time.

Guanine supramolecular architectures as 2D scaffolds

G equipped with ribose in N(9) position *via* a β-N(9)-glycosidic bond, with lipophilic chains appear as ideal building unit for the construction of complex suprastructures with a controlled rigidity, therefore opening perspectives towards their future use for scaffolding, i.e. to place functional groups in pre-designed positions.[32] Balancing the functionalities of single moieties in a supramolecular assembly represents an adaptable method for generating distinct

polymeric architectures with programmed conformations and tailored properties. In this context, we have designed a guanosine derivative ***G15***, bearing a terthiophene moiety linked to the sugar unit. Indeed, oligo- and poly-thiophenes are part of the most studied structures in organic electronics, because of their interesting optical and electronic properties, with application as active material in field-effect transistors and photovoltaic diodes noticeably. We have shown that this guanosine-terthiophene derivative can form (in solution) different types of H-bonded supramolecular architectures depending on the experimental conditions: the reversible inter-conversion fuelled by potassium ion complexation/release allows the switching between ribbons and ***G15***$_4$ self-assemblies, thus allowing us to modify the inter-oligothiophene interactions by chemical stimuli. STM and Atomic Force Microscopy (AFM) characterization showed that these molecules self-assemble into highly ordered architectures on surfaces (graphite or mica). By combining STM imaging with molecular modelling simulations, it was shown that the highly-directional structures arise from self-assembly in extended, parallel G-ribbons B characterized by a pairing N(1)-H···O(6) and N(2)-H···N(3). When adsorbed on HOPG these ribbons have been found to extend over the micrometers scale, as observed by AFM imaging of dry films. This is in contrast with previous results on alkylated guan(os)ine derivatives,[22] which showed on graphite another ribbons type, i.e. *G*-ribbons A. This difference can be explained by the fact that the guanosine-terthiophene derivative possess only one alkyl group (while guanosine ***G13*** and ***G14*** derivatives previously studied were doubly alkylated) and one acetonide group on the sugar unit (with a methyl pointing perpendicularly to the molecule main plane), both leading to several restrictions that favour the formation of a different H-bonding network. Molecular modelling suggests the formation of H-bonds between guanosine N(2)-H and the ribose of the adjacent molecule, while the spacing between ribbons is dictated by the partial interdigitation of terthiophene-alkyl groups. Indeed, this self-assembly governed by the formation of H-bonds between

guanosines dictates the spatial localization of oligothiophenes, which constitute an elegant strategy to fabricating prototypes of supramolecular nanowires for organic electronics.

2.1.2 Monocomponent self-assembly of cytosine (C), adenine (A), thymine (T) and uracil (U)

Cytosine (C)

Numerous studies of cytosine (C) assemblies on the Au (111) surface have been performed at the solid/liquid interface.[33] C molecules deposited under well-controlled UHV conditions were observed to form one-dimensional chains (or filaments) on a Cu (111) surface.[34] However, the molecular resolution in the images is rather poor, and the origin of the filamentary structure is not clear. Only in 2008 Besenbacher and co-workers demonstrated that C self-assembles into glass-like structures, whose arrangement depends on the surface coverage.[35] Among various self-assembly patterns of ***C1*** molecules (**Figure 7a**), two ribbon-like architectures, characterized by N(4)-H···O(2) and N(1)-H···N(3) or N(4)-H···N(3) and N(1)-H···O(2) pairing and hexameric macrocycle formed *via* N(4)-H···O(2), N(1)-H···N(3) and C(5)-H···N(3) H-bonds were observed (Figure 7b).

Adenine (A)

Formation of self-assembled structures of adenine (A, Figure 7a) has been extensively studied in the last two decades. When deposited onto the Cu(110) surface,[36] one-dimensional A chains are found to coexist with two-dimensional hexagonal A networks[36a] similar in shape with those observed on other substrates such as MoS_2[37] and graphite.[38] Alongside a hexagonal structure also a double-chain A structure was observed, at much lower deposition rate,[36a] and this structure has not been reported on other surfaces. Self-assembly of A molecules can be extremely complex: a total of 21 possible A dimers have been reported.[39] Therefore there may exist several competing structures that look similar in observed STM

images.[40] Nonetheless, it has been shown recently,[41] that on Au(111) surface ***A1*** (Figure 7c) molecules form highly-ordered ***A1*** 2D supramolecular network stabilized by N(9)-H···N(3), N(6)-H···N(7) and N(6)-H···N(1) H-bonds.

Thymine (T)

There are only few reports dealing with homo-assembly of thymine (T, Figure 7a) on solid substrates.[42] In 2007 Besenbacher and co-workers,[42a] reported on the adsorption of T molecules on the Au(111) surface. In particular, it has been demonstrated that ***T1*** molecules self-assemble into 1D filaments stabilized by two different pairing motifs, i.e. motif A: (Figure 7d) N(3)-H···O(2), N(3)-H···O(4) and C(6)-H···O(4), and motif B (Figure 7d): N(3)-H···O(2) and N(1)-H···O(4).

Recently, De Feyter and co-workers studied the impact of thymine[43] and thymidine[44] on the self-assembly of achiral and chiral oligo-(*p*-phenylenevinylene) (OPV) derivatives at the solid/liquid interface. In particular, the thymine-induced pattern transformation of OPV derivatives from rosettes to dimers was observed.[43] As such, the OPV derivatives ‘‘sense’’ the presence of thymine, while achiral OPV derivatives are more ‘‘sensitive’’ than chiral ones to the presence of thymine, i.e. a transition from rosettes to dimers happens at a smaller thymine to OPV ratio for the achiral derivatives. Quite unexpectedly, surface-confined supramolecular diastereomers were formed in the case of co-adsorption of achiral thymine with an enantiopure OPV derivative, leading to reduction of the degree of surface chirality. This adds to the complexity of multicomponent self-assembly, and provides a way to tune surface chirality, as it shows that it is not necessary to add the optical antipode molecule to have an impact on surface chirality.

Uracil (U)

Uracil (U) is an RNA base, and plays an important role in biological interactions, governing information transport and catalytic functions.[45] The adsorption of *U* molecules on various atomically flat substrates, e.g. Au(111),[42d, 46] Au (100),[46a, 47] Ag(111),[48] Cu (111),[48b] MoS_2,[49] HOPG[49-50] and Si (100),[51] has been extensively explored in the last two decades. In particular, Barth and co-workers[48b] showed that ***U1*** molecules (Figure 7a) adsorb flat on Ag(111), and form close-pack 2D islands. The self-assembly of ***U1*** is driven by formation of H-bonded dimers, characterized by N(3)-H···O(4) pairing, however the two-dimensional order of ***U1*** dimers is relatively poor. Interestingly, ***U1*** deprotonates at the N(3) site upon adsorption on Cu(111), and forms tiara-like structures (Figure 7e).

2.1.3 Versatile H-bonding motifs through NB-pairing

The Watson-Crick motif (Figure 1b), found in a range of structures including DNA and/or RNA motifs, is the most widely recognized H-bonding interaction in Nature. This canonical motif is defined by the pairing of guanine and cytosine (G•C) and adenine with either thymine or uracil (A•T (U)). G•C couple is stabilized by the three-point H-bonding interactions, while A•T and/or A•U dimers contain a two-point H-bonding mode. Even though the Watson-Crick mode of pairing is ubiquitous in natural systems, other H-bonding motifs are available and expand the possibility of design of different structural networks. For example, special attention needs to be paid to the Hoogsteen mode of bonding.[52] Alongside with Hoogsteen, other non-traditional base pairs are found in various DNA and RNA constructs, and include reverse Watson-Crick and reverse Hoogsteen motifs. Numerous examples of base pairing between various NBs under UHV[53] and at the solid/liquid interface,[54] have been reported. Those include 2D supramolecular assemblies based on: G•C classical Watson-Crick pairing,[53a] G•U Wobble motif,[54b] A•C Hoogsteen interactions[53a] and A•T Watson-Crick[54c] and reverse Hoogsteen patterns.[54a]

2.2. Self-assembly of unnatural nucleobases in 2D

2.2.1. Xanthine (X)

Unnatural NBs are extremely interesting from chemical and biological points of view, since the play an significant role in biology.[55] In particular, xanthine[56] (X, **Figure 8a**) and its N(9)- derivatives play a key role in a different metabolic pathways. The self-assembly of ***X1*** (Figure 8a) molecules on solid surfaces has been studied by STM under UHV,[57] and it was found that xanthine self-assembles into two extended homochiral networks tiled by two types of di-pentamer units stabilized by N(1)-H···O(2) and N(7)-H···O(6) or N(1)-H···O(6) and N(7)-H···O(2) intermolecular double H-bonds (Figure 8b).

In the collaboration with the groups of Kovács, we have explored the self-assembly at the graphite/solution interface of X exposing in N(3)-position alkyl side chains with altered lengths, i.e. -CH_3 (***X2***, Figure 8a) and –$C_{18}H_{37}$ (***X3***, Figure 8a).[58] We found, that the changes in the length of the alkyl side chains did not affected the supramolecular packing of X in 2D. ***X2*** and ***X3*** self-assembled into linear H-bonded ribbons through N(1)-H···O(2) and N(7)-H···O(6) pairings (Figure 8c and 8d). s

In the framework of a collaboration with Kovács, our group has also exploited acceptor–donor–acceptor (ADA)/donor–acceptor–donor (DAD) H-bonding to steer the formation of multi-component supramolecular structures capable of forming 2D porous networks as a result of secondary non-covalent interactions.[59] In particular, we focused our attention on the recognition process between ***X3*** molecules and 1,3,5-triazine-2,4,6-triamine (melamine, M), which resulted in the formation of M***X3***$_3$ entities through complementary H-bonding binding sites, which are further reinforced by weak C(8)-H···N(9) interactions. The possibility of generating multi component 2D porous arrays at surfaces and interfaces through the use of secondary interactions is of general interest for the formation of 2D scaffolds and offers an improved control over the supramolecular architecture, which can result in to improving the properties of the materials.

2.2.2. Isocytosine (iC)

Recently,[60] we have investigated at the graphite/solution interface the molecular physisorption of isocytosines (iC) functionalized in the C(6)-position with a phenyl ring decorated with different alkoxy side chains, i.e. 6-[4-(octyloxy)phenyl]isocytosine (***iC1***) and 6-[3,4,5-(triethoxy)phenyl] isocytosine (***iC2***) (Figure 8e). By aiming at exploring the effect of the distribution of nine atoms in different fashion, we functionalized iC one octyloxy or three ethoxy or units, which results in the subtle change of geometric properties of isocytosine derivatives. Such a modification, which primarily can be associated to a change in steric hindrance, can in turn influence the process of self-assembly at the solid/liquid interface, affecting the supramolecular order at surfaces.[61] Derivative ***iC1,*** equipped OC_8H_{17} chain, forms linear H-bonded ribbons through the N(1)-H···O(4), N(2)-H···O(4) and N(2)-H···N(3) pairing (Figure 8f). Differently, as a result of N(1)-H···O(4) and N(2)-H···N(3) H-pairing, hexameric supramolecular macrocycles are formed by derivative ***iC2***. Overall, by controlling geometrical constraints it is possible to steer the molecular assembly towards the generation of either ribbons or cyclic hexamers.

3. Conclusion

While many elegant NB assemblies containing two-dimensional supramolecular structures have been fabricated on surfaces and interfaces, it is nonetheless clear that much remains to be done. When confronted with NB interactions majority of scientists will automatically think of the Watson–Crick base pairing, however over the past years researchers from a variety of different disciplines have shown that NBs are much more versatile with their bonding behavior. In this *Review*, we have shown, that with the help of scanning tunnelling microscopy, the adsorption of natural and unnatural NBs can be visualized with the sub-nanometer precision. By identifying the structure of the two-dimensional NB patterns, detailed information about the recognition process on the surfaces and interfaces has been

obtained. We have provided evidence here for these early attempts of pre-programming self-assembly of NB, being a topic extremely relevant for many fields of research.

Self-assembly at surfaces and interfaces is with no doubt the most systematically studied field towards the *bottom-up* fabrication of controlled supramolecular architectures. Therefore, to fully control the self-assembly of NB-based building blocks in 2D, i.e. at surfaces and interfaces, theoretical approaches have to be exploited. Recent advances in the theoretical framework as well as the long series of successful results reported in this *Review* provide unprecedented insights into the first principles of NB self-organization in 2D and will hopefully help the development of computer algorithms and methods to predict their surface-assisted self-assembly.[62]

Over the past decades, chemists have effectively reproduced the exquisite specificity of biomolecular interactions and their adaptive nature.[63] Nonetheless, engineering multiple specific interactions in abiotic systems remains challenging. DNA retains its position as the best medium to create orthogonal, isoenergetic interactions based on the recognition between NBs. [64] Numerous exciting examples and approaches have been reported in the last decade, and include Rothemund's origami,[65] Gothlef and co-workers' DNA box,[66] Seeman[67] and Dietz[68] DNA-based nanopatterns, to name a few. The capability to further understand how nucleobases interact both with themselves as well as with other molecules will not only allow the fabrication of hybrid systems which can bind more specifically to targeted DNA sequences but will also allow us to gain valuable insights on how we might be able to harness these interesting biological molecules to construct more and more complex multifunctional 2D and 3D nanostructures and (nano)materials.

Acknowledgements

We are grateful to our collaborators, the late Professor Gian Piero Spada (Bologna), and Lajos Kovács (Szeged) for the joint endeavour on nucleobases. This work was supported by the European Community through the project EC FP7 ICT-MOLARNET (318516) and the

European Research Council project SUPRAFUNCTION (GA-257305), the Agence Nationale de la Recherche through the LabEx project Chemistry of Complex Systems (ANR-10-LABX-0026_CSC) and the International Center for Frontier Research in Chemistry.

Received: ((will be filled in by the editorial staff))
Revised: ((will be filled in by the editorial staff))
Published online: ((will be filled in by the editorial staff))

[1] a) L. Brunsveld, B. J. B. Folmer, E. W. Meijer, R. P. Sijbesma, *Chem. Rev.* **2001**, *101*, 4071-4097; b) J. A. A. W. Elemans, A. E. Rowan, R. J. M. Nolte, *J. Mater. Chem.* **2003**, *13*, 2661-2670; c) T. F. A. Greef, E. W. Meijer, *Nature* **2008**, *453*, 171-173; d) J.-M. Lehn, *Supramolecular chemistry: concepts and perspectives*, VCH New York, **1995**; e) J.-M. Lehn, *Polym. Int.* **2002**, *51*, 825-839; f) J. H. van Esch, B. L. Feringa, *Angew. Chem. Int. Ed.* **2000**, *39*, 2263-2266.

[2] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737-738.

[3] W. Saenger, *Principles of nucleic acid structure*, Springer-Verlag, New York, **1983**.

[4] a) S. Sivakova, S. J. Rowan, *Chem. Soc. Rev.* **2005**, *34*, 9-21; b) J. L. Sessler, C. M. Lawrence, J. Jayawickramarajah, *Chem. Soc. Rev.* **2007**, *36*, 314-325; c) J. L. Sessler, J. Jayawickramarajah, *Chem. Commun.* **2005**, 1939-1949; d) S. A. Benner, *Acc. Chem. Res.* **2004**, *37*, 784-797; e) R. E. A. Kelly, L. N. Kantorovich, *J. Mater. Chem.* **2006**, *16*, 1894-1905; f) L. Liu, D. Xia, L. H. Klausen, M. D. Dong, *Int. J. Mol. Sci.* **2014**, *15*, 1901-1914.

[5] a) G. Binnig, H. Rohrer, C. Gerber, E. Weibel, *Phys. Rev. Lett.* **1982**, *49*, 57-61; b) H. Rohrer, *Proc. Nat. Acad. Sci. USA* **1987**, *84*, 4666-4666.

[6] a) J. P. Rabe, S. Buchholz, *Science* **1991**, *253*, 424-427; b) S. De Feyter, A. Gesquiere, M. M. Abdel-Mottaleb, P. C. M. Grim, F. C. De Schryver, C. Meiners, M. Sieffert, S. Valiyaveettil, K. Müllen, *Acc. Chem. Res.* **2000**, *33*, 520-531; c) A. M. Jackson, J. W. Myerson, F. Stellacci, *Nat. Mater.* **2004**, *3*, 330-336; d) A. Cadeddu, A. Ciesielski, T. El Malah, S. Hecht, P. Samorì, *Chem. Commun.* **2011**, *47*, 10578-10580; e) A. Ciesielski, A. Cadeddu, C. A. Palma, A. Gorczynski, V. Patroniak, M. Cecchini, P. Samorì, *Nanoscale* **2011**, *3*, 4125-4129; f) J. M. MacLeod, O. Ivasenko, D. F. Perepichka, F. Rosei, *Nanotechnology* **2007**, *18*, 424031-424040; g) A. Ciesielski, M. El Garah, S. Haar, P. Kovaricek, J. M. Lehn, P. Samorì, *Nat. Chem.* **2014**, *6*, 1017-1023; h) A. Ciesielski, C.-A. Palma, M. Bonini, P. Samorì, *Adv. Mater.* **2010**, *22*, 3506-3520; i) A. Ciesielski, L. Piot, P. Samorì, A. Jouaiti, M. W. Hosseini, *Adv. Mater.* **2009**, *21*, 1131-1136; j) A. Ciesielski, P. Samorì, *Nanoscale* **2011**, *3*, 1397-1410; k) A. Ciesielski, G. Schaeffer, A. Petitjean, J.-M. Lehn, P. Samorì, *Angew. Chem. Int. Ed.* **2009**, *48*, 2039-2043; l) M. E. Canas-Ventura, W. Xiao, D. Wasserfallen, K. Müllen, H. Brune, J. V. Barth, R. Fasel, *Angew. Chem. Int. Ed.* **2007**, *46*, 1814-1818; m) S. Stepanow, M. Lingenfelder, A. Dmitriev, H. Spillmann, E. Delvigne, N. Lin, X. Deng, C. Cai, J. V. Barth, K. Kern, *Nat. Mater.* **2004**, *3*, 229-233.

[7] R. Lazzaroni, A. Calderone, J. L. Bredas, J. P. Rabe, *J. Chem. Phys.* **1997**, *107*, 99-105.

[8] P. Samorì, K. Müllen, H. P. Rabe, *Adv. Mater.* **2004**, *16*, 1761-1765.

[9] L. Piot, R. M. Meudtner, T. El Malah, S. Hecht, P. Samorì, *Chem. Eur. J.* **2009**, *15*, 4788-4792.

[10] a) M. Surin, P. Samorì, A. Jouaiti, N. Kyritsakas, M. W. Hosseini, *Angew. Chem. Int. Ed.* **2007**, *46*, 245-249; b) M. El Garah, A. Ciesielski, N. Marets, V. Bulach, M. W. Hosseini, P. Samorì, *Chem. Commun.* **2014**, *50*, 12250-12253.

[11] a) B. Venkataraman, J. J. Breen, G. W. Flynn, *J. Phys. Chem.* **1995**, *99*, 6608-6619; b) C. J. Li, Q. D. Zeng, C. Wang, L. J. Wan, S. L. Xu, C. R. Wang, C. L. Bai, *J. Phys. Chem. B* **2003**, *107*, 747-750; c) M. Lackinger, S. Griessl, T. Markert, F. Jamitzky, W. M. Heckl, *J. Phys. Chem. B* **2004**, *108*, 13652-13655; d) M. Lackinger, S. Griessl, W. A. Heckl, M. Hietschold, G. W. Flynn, *Langmuir* **2005**, *21*, 4984-4988; e) L. Kampschulte, M. Lackinger, A. K. Maier, R. S. K. Kishore, S. Griessl, M. Schmittel, W. M. Heckl, *J. Phys. Chem. B* **2006**, *110*, 10829-10836; f) W. Mamdouh, H. Uji-i, J. S. Ladislaw, A. E. Dulcey, V. Percec, F. C. De Schryver, S. De Feyter, *J. Am. Chem. Soc.* **2006**, *128*, 317-325; g) K. G. Nath, O. Ivasenko, J. A. Miwa, H. Dang, J. D. Wuest, A. Nanci, D. F. Perepichka, F. Rosei, *J. Am. Chem. Soc.* **2006**, *128*, 4212-4213.

[12] R. Gutzler, L. Cardenas, F. Rosei, *Chem. Sci.* **2011**, *2*, 2290-2300.

[13] P. Samorì, N. Severin, K. Müllen, J. P. Rabe, *Adv. Mater.* **2000**, *12*, 579-582.

[14] a) J. T. Davis, *Angew. Chem. Int. Ed.* **2004**, *43*, 668-698; b) N. Sreenivasachary, J.-M. Lehn, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5938-5943; c) J. T. Davis, G. P. Spada, *Chem. Soc. Rev.* **2007**, *36*, 296-313; d) S. Lena, S. Masiero, S. Pieraccini, G. P. Spada, *Chem. Eur. J.* **2009**, *15*, 7792-7806; e) W. Fritzsche, L. Spindler, *Guanine quartets: structure and application*, Royal Society of Chemistry, Cambridge, **2012**.

[15] M. W. Heckl, D. P. E. Smith, G. Binnig, H. Klagges, T. W. Hansch, J. Maddocks, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8003-8005.

[16] R. Otero, M. Schock, L. M. Molina, E. Laegsgaard, I. Stensgaard, B. Hammer, F. Besenbacher, *Angew. Chem. Int. Ed.* **2005**, *44*, 2270-2275.

[17] W. Xu, R. E. A. Kelly, H. Gersen, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2009**, *5*, 1952-1956.

[18] A. P. Wolffe, M. A. Matzke, *Science* **1999**, *286*, 481-486.

[19] I. Bald, Y. G. Wang, M. D. Dong, C. B. Rosen, J. B. Ravnsbaek, G. L. Zhuang, K. V. Gothelf, J. G. Wang, F. Besenbacher, *Small* **2011**, *7*, 939-949.

[20] L. K. Wang, H. H. Kong, C. Zhang, Q. Sun, L. L. Cai, Q. G. Tan, F. Besenbacher, W. Xu, *ACS Nano* **2014**, *8*, 11799-11805.

[21] A. Ciesielski, R. Perone, S. Pieraccini, G. P. Spada, P. Samorì, *Chem. Commun.* **2010**, *46*, 4493-4495.

[22] S. Lena, G. Brancolini, G. Gottarelli, P. Mariani, S. Masiero, A. Venturini, V. Palermo, O. Pandoli, S. Pieraccini, P. Samorì, G. P. Spada, *Chem. Eur. J.* **2007**, *13*, 3757-3764.

[23] G. Gottarelli, S. Masiero, E. Mezzina, S. Pieraccini, J. P. Rabe, P. Samorì, G. P. Spada, *Chem. Eur. J.* **2000**, *6*, 3242-3248.

[24] U. Thewalt, C. E. Bugg, R. Marsh, *Acta Crystallogr. Sect. B: Structural Crystallography and Crystal Chemistry* **1970**, *26*, 1089-1101.

[25] G. Maruccio, P. Visconti, V. Arima, S. D'Amico, A. Blasco, E. D'Amone, R. Cingolani, R. Rinaldi, S. Masiero, T. Giorgi, G. Gottarelli, *Nano Lett.* **2003**, *3*, 479-483.

[26] T. Giorgi, S. Lena, P. Mariani, M. A. Cremonini, S. Masiero, S. Pieraccini, J. P. Rabe, P. Samorì, G. P. Spada, G. Gottarelli, *J. Am. Chem. Soc.* **2003**, *125*, 14741-14749.

[27] a) M. D. Topal, J. R. Fresco, *Nature* **1976**, *263*, 285-289; b) M. Y. Choi, R. E. Miller, *J. Am. Chem. Soc.* **2006**, *128*, 7320-7328.

[28] P.-O. Löwdin, *Rev. Mod. Phys.* **1963**, *35*, 724-732.

[29] H. H. Kong, Q. Sun, L. K. Wang, Q. G. Tan, C. Zhang, K. Sheng, W. Xu, *ACS Nano* **2014**, *8*, 1804-1808.

[30] A. Ciesielski, S. Lena, S. Masiero, G. P. Spada, P. Samorì, *Angew. Chem. Int. Ed.* **2010**, *49*, 1963-1966.

[31] J.-M. Lehn, *Prog. Polym. Sci.* **2005**, *30*, 814-831.

[32] G. P. Spada, S. Lena, S. Masiero, S. Pieraccini, M. Surin, P. Samorì, *Adv. Mater.* **2008**, *20*, 2433-2438.

[33] a) T. Wandlowski, D. Lampner, S. M. Lindsay, *J. Electroanal. Chem.* **1996**, *404*, 215-226; b) M. Komiyama, M. Gu, T. Shimaguchi, H. M. Wu, T. Okada, *Appl. Phys. A: Mater. Sci. Process.* **1998**, *66*, S635-S637.
[34] a) H. Tanaka, T. Nakagawa, T. Kawai, *Surf. Sci.* **1996**, *364*, L575-L579; b) T. Kawai, H. Tanaka, T. Nakagawa, *Surf. Sci.* **1997**, *386*, 124-136; c) M. Furukawa, H. Tanaka, T. Kawai, *J. Chem. Phys.* **2001**, *115*, 3419-3423.
[35] a) R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Science* **2008**, *319*, 312-315; b) R. E. A. Kelly, M. Lukas, L. N. Kantorovich, R. Otero, W. Xu, M. Mura, E. Laegsgaard, I. Stensgaard, F. Besenbacher, *J. Chem. Phys.* **2008**, *129*, 184707-184782.
[36] a) M. Furukawa, H. Tanaka, T. Kawai, *Surf. Sci.* **2000**, *445*, 1-10; b) Q. Chen, D. J. Frankel, N. V. Richardson, *Langmuir* **2002**, *18*, 3219-3225.
[37] S. J. Sowerby, W. M. Heckl, G. B. Petersen, *J. Mol. Evol.* **1996**, *43*, 419-424.
[38] a) M. Edelwirth, J. Freund, S. J. Sowerby, W. M. Heckl, *Surf. Sci.* **1998**, *417*, 201-209; b) J. E. Freund, M. Edelwirth, P. Krobel, W. M. Heckl, *Phys. Rev. B* **1997**, *55*, 5394-5397; c) N. J. Tao, Z. Shi, *J. Phys. Chem.* **1994**, *98*, 1464-1471.
[39] R. E. A. Kelly, Y. J. Lee, L. N. Kantorovich, *J. Phys. Chem. B* **2005**, *109*, 11933-11939.
[40] L. M. A. Perdigao, P. A. Staniec, N. R. Champness, R. E. A. Kelly, L. N. Kantorovich, P. H. Beton, *Phys. Rev. B* **2006**, *73*, 195423-195430.
[41] a) R. E. A. Kelly, W. Xu, M. Lukas, R. Otero, M. Mura, Y. J. Lee, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2008**, *4*, 1494-1500; b) M. Lukas, R. E. A. Kelly, L. N. Kantorovich, R. Otero, W. Xu, E. Laegsgaard, I. Stensgaard, F. Besenbacher, *J. Chem. Phys.* **2009**, *130*, 024705-024711.
[42] a) W. Xu, R. E. A. Kelly, R. Otero, M. Schock, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2007**, *3*, 2011-2014; b) A. Chatterjee, L. Zhang, K. T. Leung, *J. Phys. Chem. C* **2013**, *117*, 14677-14683; c) B. Roelfs, E. Bunge, C. Schroter, T. Solomun, H. Meyer, R. J. Nichols, H. Baumgartel, *J. Phys. Chem. B* **1997**, *101*, 754-765; d) W. H. Li, W. Haiss, S. Floate, R. J. Nichols, *Langmuir* **1999**, *15*, 4875-4883.
[43] Z. Guo, I. De Cat, B. Van Averbeke, J. Lin, G. Wang, H. Xu, R. Lazzaroni, D. Beljonne, A. P. H. J. Schenning, S. De Feyter, *Chem. Commun.* **2014**, *50*, 11903-11906.
[44] a) Z. Guo, I. De Cat, B. Van Averbeke, E. Ghijsens, J. Lin, H. Xu, G. Wang, F. J. M. Hoeben, Z. Tomovic, R. Lazzaroni, D. Beljonne, E. W. Meijer, A. P. H. J. Schenning, S. De Feyter, *J. Am. Chem. Soc.* **2013**, *135*, 9811-9819; b) Z. X. Guo, I. De Cat, B. Van Averbeke, J. B. Lin, G. J. Wang, H. Xu, R. Lazzaroni, D. Beljonne, E. W. Meijer, A. P. H. J. Schenning, S. De Feyter, *J. Am. Chem. Soc.* **2011**, *133*, 17764-17771.
[45] B. G. Vertessy, J. Toth, *Acc. Chem. Res.* **2009**, *42*, 97-106.
[46] a) T. Dretschkow, A. S. Dakkouri, T. Wandlowski, *Langmuir* **1997**, *13*, 2843-2856; b) F. Cunha, F. Nart, *J. Braz. Chem. Soc.* **2001**, *12*, 715-721; c) F. Cunha, E. Sa, F. Nart, *Surf. Sci.* **2001**, *480*, L383-L388.
[47] T. Dretschkow, T. Wandlowski, *Electrochim. Acta* **1998**, *43*, 2991-3006.
[48] a) M. Cavallini, G. Aloisi, M. Bracali, R. Guidelli, *J. Electroanal. Chem.* **1998**, *444*, 75-81; b) A. C. Papageorgiou, S. Fischer, J. Reichert, K. Diller, F. Blobner, F. Klappenberger, F. Allegretti, A. P. Seitsonen, J. V. Barth, *ACS Nano* **2012**, *6*, 2477-2486.
[49] S. J. Sowerby, M. Edelwirth, W. M. Heckl, *Appl. Phys. A* **1998**, *66*, S649-S653.
[50] S. J. Sowerby, G. B. Petersen, *J. Electroanal. Chem.* **1997**, *433*, 85-90.
[51] A. Lopez, Q. Chen, N. V. Richardson, *Surf. Interface Anal.* **2002**, *33*, 441-446.
[52] K. Hoogsteen, *Acta Crystallogr.* **1963**, *16*, 907-916.

[53] a) R. Otero, W. Xu, M. Lukas, R. E. A. Kelly, E. Laegsgaard, I. Stensgaard, J. Kjems, L. N. Kantorovich, F. Besenbacher, *Angew. Chem. Int. Ed.* **2008**, *47*, 9673-9676; b) D. Y. Zhong, T. Blomker, C. Muck-Lichtenfeld, H. M. Zhang, G. Kehr, G. Erker, H. Fuchs, L. F. Chi, *Small* **2014**, *10*, 265-270.

[54] a) W. Mamdouh, M. D. Dong, S. L. Xu, E. Rauls, F. Besenbacher, *J. Am. Chem. Soc.* **2006**, *128*, 13305-13311; b) W. Mamdouh, R. E. A. Kelly, M. D. Dong, L. N. Kantorovich, F. Besenbacher, *J. Am. Chem. Soc.* **2008**, *130*, 695-702; c) A. G. Slater, Y. Hu, L. X. Yang, S. P. Argent, W. Lewis, M. O. Blunt, N. R. Champness, *Chem. Sci.* **2015**, *6*, 1562-1569; d) S. L. Xu, M. D. Dong, E. Rauls, R. Otero, T. R. Linderoth, F. Besenbacher, *Nano Lett.* **2006**, *6*, 1434-1438.

[55] D. T. Hurst, *An introduction to the chemistry and biochemistry of pyrimidines, purines, and pteridines*, John Wiley & Sons, New York, **1980**.

[56] a) S. Miyakawa, H. J. Cleaves, S. L. Miller, *Origins Life Evol. Biosphere* **2002**, *32*, 209-218; b) E. Kulikowska, B. Kierdaszuk, D. Shugar, *Acta Biochim. Pol.* **2004**, *51*, 493-531; c) I. Biaggioni, S. Paul, A. Puckett, C. Arzubiaga, *J. Pharmacol. Exp. Ther.* **1991**, *258*, 588-593.

[57] M. Yu, J. G. Wang, M. Mura, Q. Q. Meng, W. Xu, H. Gersen, E. Laegsgaard, I. Stensgaard, R. E. A. Kelly, J. Kjems, T. R. Linderoth, L. N. Kantorovich, F. Besenbacher, *ACS Nano* **2011**, *5*, 6651-6660.

[58] A. Ciesielski, S. Haar, A. Benyei, G. Paragi, C. F. Guerra, F. M. Bickelhaupt, S. Masiero, J. Szolomajer, P. Samori, G. P. Spada, L. Kovacs, *Langmuir* **2013**, *29*, 7283-7290.

[59] A. Ciesielski, S. Haar, G. Paragi, Z. Kupihár, Z. Kele, S. Masiero, C. F. Guerra, F. M. Bickelhaupt, G. P. Spada, L. Kovács, P. Samori, *Phys. Chem. Chem. Phys.* **2013**, *15*, 12442-12446.

[60] A. Ciesielski, S. Colella, L. Zalewski, B. Bruchmann, P. Samorì, *CrystEngComm* **2011**, *13*, 5535-5537.

[61] a) S. Furukawa, H. Uji-i, K. Tahara, T. Ichikawa, M. Sonoda, F. C. De Schryver, Y. Tobe, S. De Feyter, *J. Am. Chem. Soc.* **2006**, *128*, 3502-3503; b) K. Tahara, S. Okuhata, J. Adisoejoso, S. B. Lei, T. Fujita, S. De Feyter, Y. Tobe, *J. Am. Chem. Soc.* **2009**, *131*, 17583-17590; c) J. A. A. W. Elemans, S. B. Lei, S. De Feyter, *Angew. Chem. Int. Ed.* **2009**, *48*, 7298-7332.

[62] C.-A. Palma, M. Cecchini, P. Samorì, *Chem. Soc. Rev.* **2012**, *41*, 3713-3730.

[63] F. A. Aldaye, H. F. Sleiman, *J. Am. Chem. Soc.* **2007**, *129*, 13376-+.

[64] a) N. C. Seeman, *Nature* **2003**, *421*, 427-431; b) F. A. Aldaye, A. L. Palmer, H. F. Sleiman, *Science* **2008**, *321*, 1795-1799; c) P. W. K. Rothemund, *Nature* **2006**, *440*, 297-302.

[65] P. W. K. Rothemund, *Nature* **2006**, *440*, 297-302.

[66] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, J. Kjems, *Nature* **2009**, *459*, 73-76.

[67] T. Wang, R. J. Sha, R. Dreyfus, M. E. Leunissen, C. Maass, D. J. Pine, P. M. Chaikin, N. C. Seeman, *Nature* **2011**, *478*, 225-228.

[68] J. P. J. Sobczak, T. G. Martin, T. Gerling, H. Dietz, *Science* **2012**, *338*, 1458-1461.

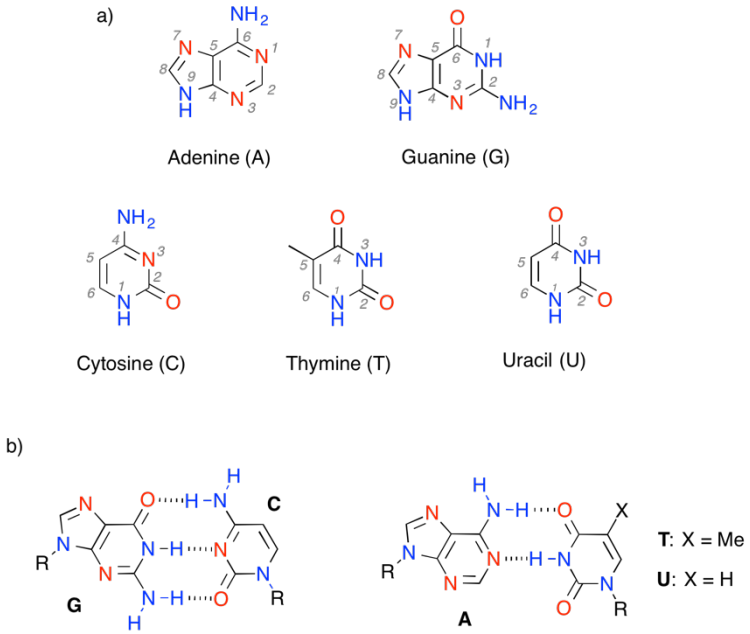

Figure 1. a) Chemical structures of guanine (*G*), cytosine (C), adenine (A), thymine (T) and uracil (U); hydrogen-bonding donor and acceptor sites are indicated in blue and red, respectively, b) the canonical Watson-Crick hydrogen-bonding G•C and A•T (U) modes.

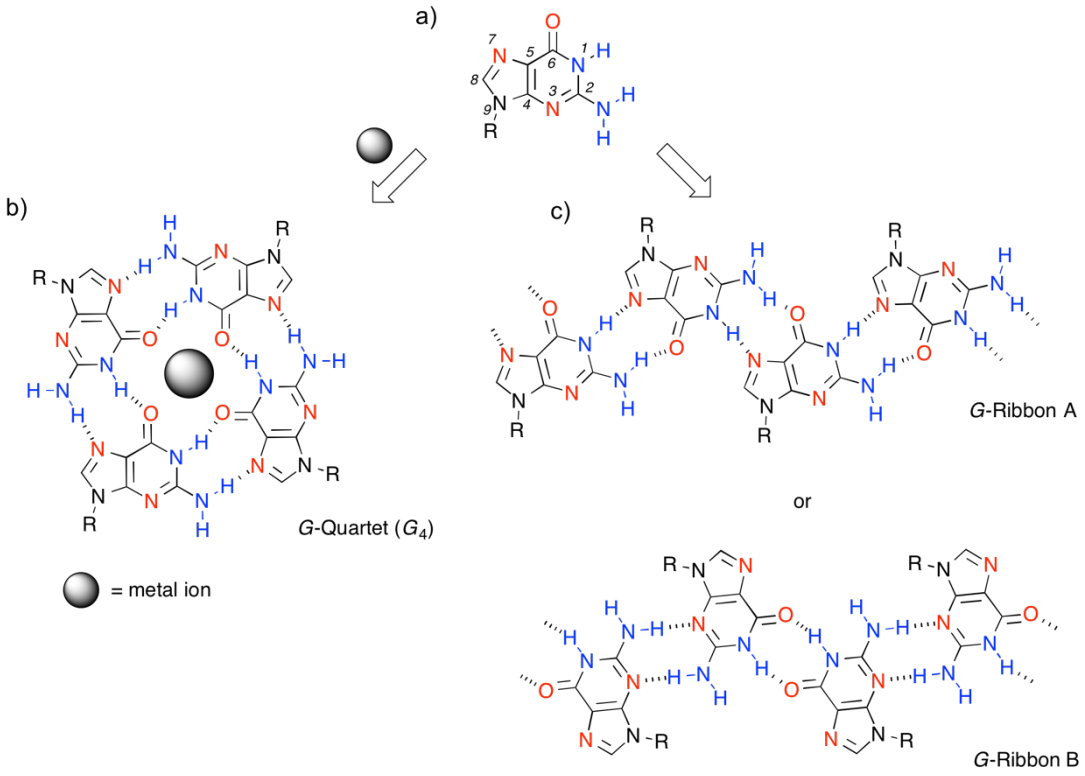

Figure 2. a) Chemical structure of guanine (G); examples of H-bonded motifs of G: b) guanine quartet (G_4) templated by the metal ion, involving N(2)-H···N(7) and N(1)-H···O(6) H-bonding and c) ribbon-like architectures G-ribbon A (H-bonding: N(2)-H···O(6) and N(1)-H···N(7)) and G-ribbon B (H-bonding: N(1)-H···O(6) and N(2)-H···N(3)). Hydrogen-bonding donor and acceptor sites are indicated in blue and red, respectively.

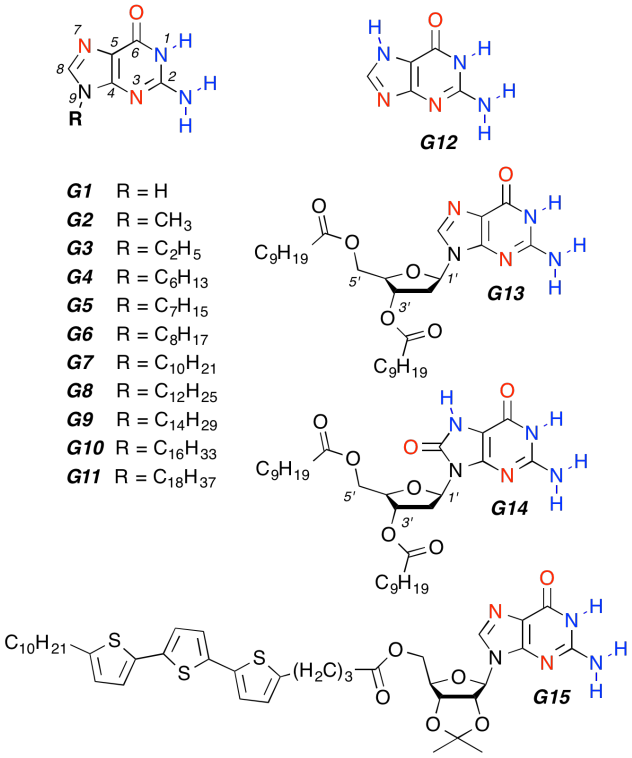

Figure 3. Chemical structures of guanine (G) derivatives, and their corresponding acronyms as used in the text. Hydrogen-bonding donor and acceptor sites are indicated in blue and red, respectively.

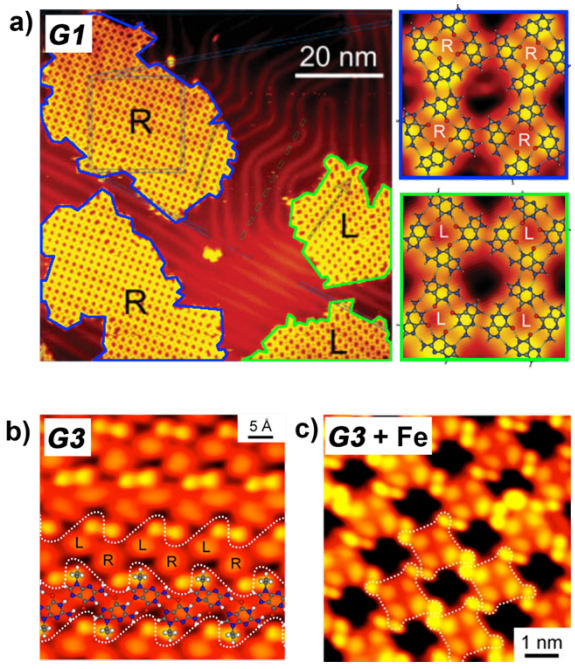

Figure 4. a) STM image containing mirror phases of enantiomerically pure R (right-handed) and L (left-handed) ***G1***$_4$-based 2D H-bonded networks self-assembled on Au(111) surface under UHV. Reproduced with permission.[17] 2009, Wiley-VCH. b) Self-assembled ***G3***-ribbon A structure formed by ***G3*** molecules on Au(111); c) STM image of ***G3***$_4$-based supramolecular architecture formed upon mixing ***G3*** with Fe ions. b-c) Reproduced with permission.[20] 2014, American Chemical Society.

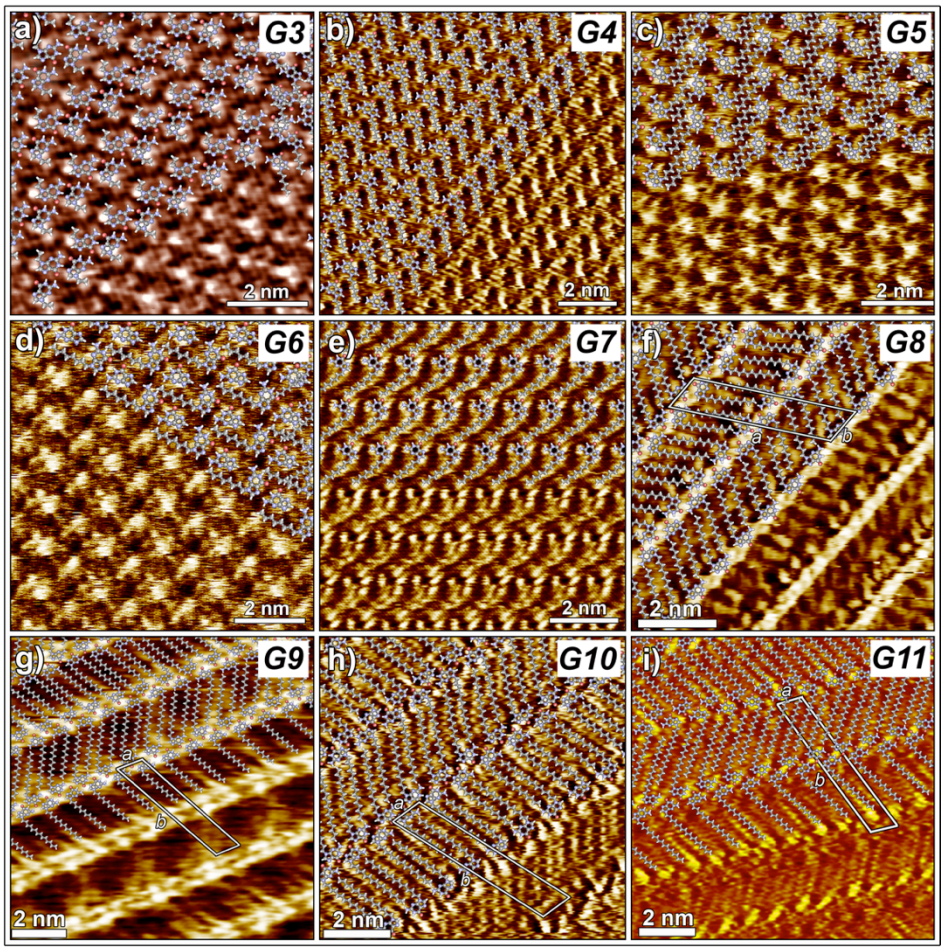

Figure 5. STM current images of monolayers of alkylated guanines showing ribbon-like (a, f, g, h and i), crystalline (b, e) and dimeric (c, d) structures formed at the HOPG/solution interface. Reproduced with permission.[21] 2010, Royal Society of Chemistry.

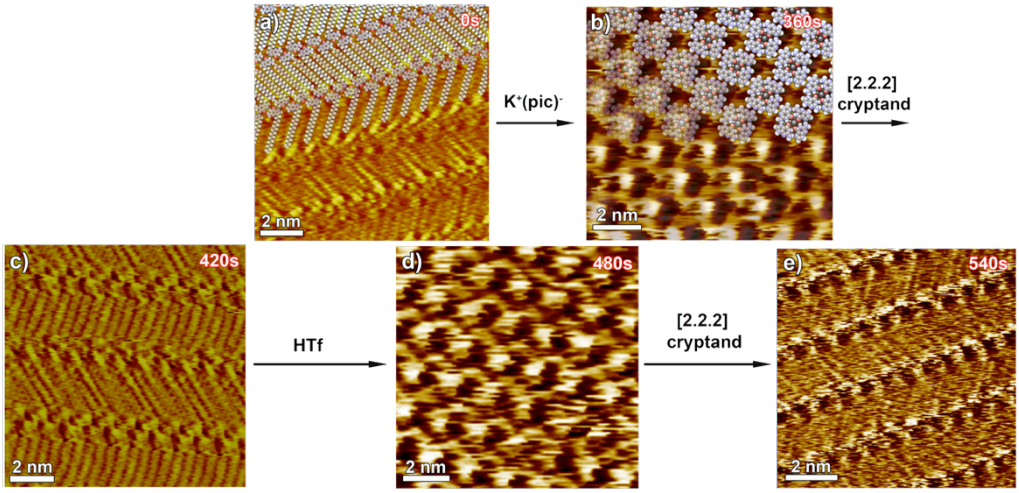

Figure 6. Consecutive STM images showing the structural evolution of a monolayer of ***G11*** over a 9 minutes time scale (time range displays in the upper right part of the images correspond to the time that was needed to reach the equilibrium after addition of reacting agents). Images (b), (d), and (f) show ***G11***-ribbon A-like structure, whereas (c) and (e) exhibit ***G11***$_4$-based architectures. Reproduced with permission.[30] 2010, Wiley-VCH.

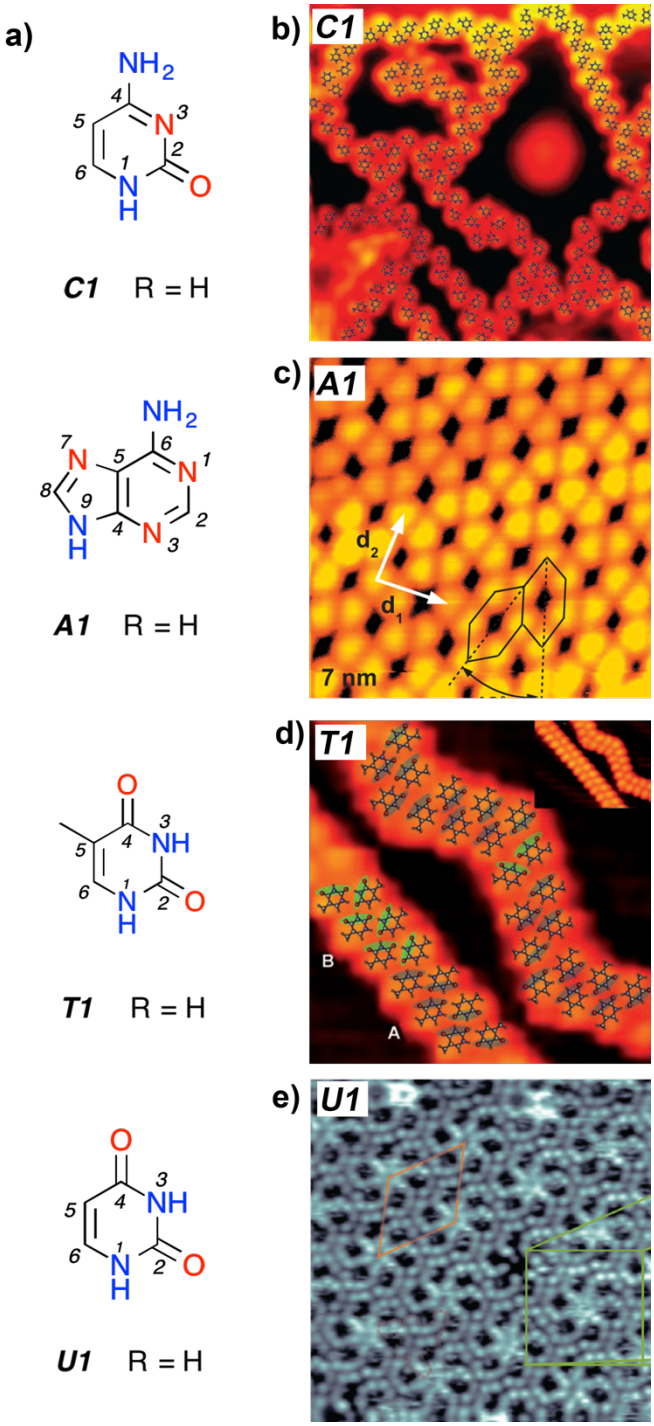

Figure 7. Chemical structures of cytosine (C), adenine (A), thymine (T) and uracil (U), and their corresponding acronyms as used in the text. Hydrogen-bonding donor and acceptor sites are indicated in blue and red, respectively. b) STM image of a "glassy state" of ***C1*** on Au(111), with an overlay illustrating several elementary H-bonded structural motifs. Reproduced with permission.[35a] 2008, American Association for the Advancement of Science. c) Self-assembled architectures formed by ***A1*** molecules on Au(111). Reproduced with permission.[41b] 2009, American Institute of Physics. d) Ribbon-like assembly of ***T1*** molecules on Au(111). Reproduced with permission.[42a] 2007, Wiley-VCH e) STM image of sub-monolayer coverage of ***U1*** on Cu(111) showing a domain of the tiara-phase. Reproduced with permission.[48b] 2012, American Chemical Society.

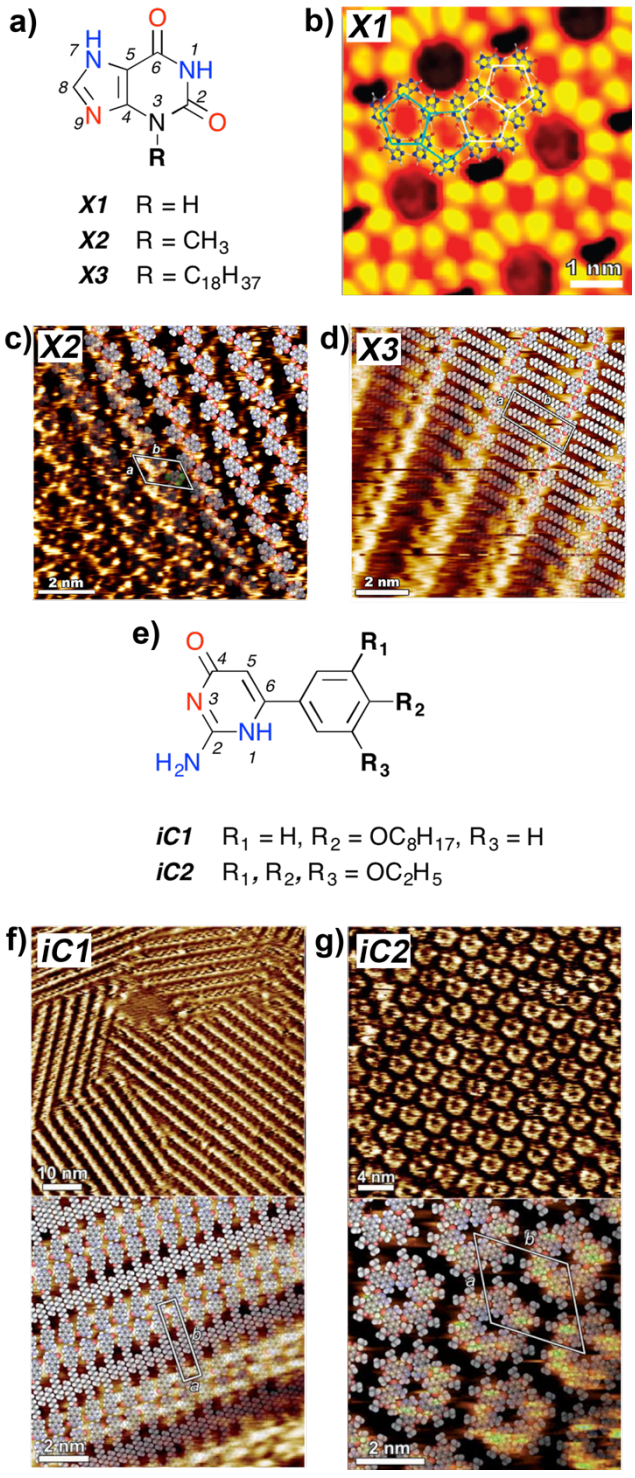

Figure 8. a) Chemical structures of xanthine (X), and its corresponding acronyms as used in the text. b) STM image of the ***X1*** sunflower structure on Au(111) showing a formation tiled by di-pentamers in two orientations, marked in white and blue pentagons. Reproduced with permission.[57] 2011, American Chemical Society. c-d) STM height images of self-assembled supramolecular H-bonded polymers of (c) ***X2*** and (d) ***X3*** at the solid/liquid interface. Reproduced with permission.[58] 2013, American Chemical Society. e) Chemical structures of isocytosine (iC), and its corresponding acronyms as used in the text. f-g) STM images showing ribbon-like supramolecular structures of ***iC1*** (f) or macrocyclic hexameric structures of ***iC2*** self-assembled at the solid–liquid interface. Reproduced with permission.[60] 2011, The Royal Society of Chemistry.

Table 1. Hydrogen bonding pairing in various 2D self-assembled motifs of nucleobases

NB	H-bonding pattern [a]	Self-assembled motif [b]	Substrate [c]
Guanine (G)	N(2)-H···O(6) and N(1)-H···N(7)	Ribbon A	Au(111)[20], HOPG[21],[23],[26]
	N(1)-H···O(6) and N(2)-H···N(3)	Ribbon B	HOPG[32]
	N(2)-H···N(7) and N(1)-H···O(6) [c]	Quartet	Au(111)[16],[17],[20], HOPG[30]
Cytosine (C)	N(4)-H···O(2) and N(1)-H···N(3)	Ribbon I	Au(111)[35]
	N(4)-H···N(3) and N(1)-H···O(2)	Ribbon II	Au(111)[35]
	N(4)-H···O(2), N(1)-H···N(3) and C(5)-H···N(3)	Hexameric macrocycle	Au(111)[35]
Adenine (A)	N(9)-H···N(3), N(6)-H···N(7) and N(6)-H···N(1)	2D network	Au(111)[41]
Thymine (T)	N(3)-H···O(2), N(3)-H···O(4) and C(6)-H···O(4)	1D filament I	Au(111)[42a]
	N(3)-H···O(2) and N(1)-H···O(4)	1D filament II	Au(111)[42a]
Uracil (U)	N(3)-H···O(4)	Dimer	Ag(111)[48b]
Xanthine (X)	N(1)-H···O(2) and N(7)-H···O(6)	Pentamer I [d]	Au(111)[57]
	N(1)-H···O(6) and N(7)-H···O(2)	Pentamer II [d]	Au(111)[57]
	N(1)-H···O(2) and N(7)-H···O(6)	Ribbon	HOPG[58]
Isocytosine (iC)	N(1)-H···O(4), N(2)-H···O(4) and N(2)-H···N(3)	Ribbon	HOPG[60]

[a)] Noteworthy, the primary type of interactions between NBs is the H-bonding. Nonetheless, depending of the solid substrate and the chemical modifications of NB structure the self-assembled motifs are stabilized by van der Waals interactions between NBs and substrate (physisorption) and van der Waals interactions between molecules, such as interdigitation between alkyl chain substituents; [b)] Nomenclature used to describe the self-assembled patterns originates from the articles cited in the row *Substrate*; [c)] In the case of guanine-based architectures visualized on Au(111) surface, secondary H-bonding pairing motif is observed, i.e. N(9)-H···N(7); [d)] Noteworthy, both pentamers co-exist in the same 2D structure.