Surface Nucleobase Pairing

DOI: 10.1002/ange.200803333

# Specificity of Watson–Crick Base Pairing on a Solid Surface Studied at the Atomic Scale**

*Roberto Otero, Wei Xu, Maya Lukas, Ross E. A. Kelly, Erik Lægsgaard, Ivan Stensgaard, Jørgen Kjems, Lev N. Kantorovich, and Flemming Besenbacher**

The fascinating double-helix structure of DNA, discovered by Watson and Crick fifty years ago,[1] provides a simple model for DNA replication based on the principle of complementary base-pairing sets of guanine (G)/cytosine (C) and adenine (A)/thymine (T). This base-pairing scheme is thought to play a crucial role in the fidelity with which DNA is replicated,[1] since it would impose an enthalpy penalty to form double-stranded DNA from error-containing strands. In the absence of polymerases in the prebiotic soup,[2,3] base pairing triggered by hydrogen bonds is thought to be the crucial factor for the recognition of nucleobases, and base pairing probably also played an important role in the polymerization of the first oligonucleotide. It has been shown that short RNA strands can act as templates that catalyze the polymerization of complementary RNA strands from activated nucleotides in solution.[4,5] However, the reaction proceeds slowly, and the replication process is relatively unfaithful.[4] Better estimations of binding energies associated with all possible interactions involved in nucleobase recognition would help to clarify the issue, but hitherto it has been difficult to design experiments in which the different contributions of hydrogen bonding, solvation energy, and hydrophobic and van der Waals interactions can be determined separately. The development of scanning probe microscopy methods has given us an invaluable tool to explore intermolecular interactions in the presence of a surface.[6,7]

In this communication, from an interplay of variable-temperature scanning tunneling microscopy (VT-STM) and DFT calculations, we investigate the feasibility of molecular recognition in the complementary G + C system, as compared with the noncomplementary A + C system, using the inert Au(111) surface as a model substrate that, while keeping the molecules adsorbed, does not restrict their mobility in the remaining two dimensions. Our results show that under extremely clean ultrahigh-vacuum (UHV) conditions both A–C and G–C hydrogen-bonded pairs are formed at room temperature, but they show different resilience towards heating. Molecular recognition thus takes place at slightly elevated temperatures above room temperature, and the difference in the thermal stabilities of G–C and A–C base pairs allows us to extract a lower bound for the difference in the hydrogen-bonding energies involved. Such information is of the utmost relevance to modeling interactions of DNA bases in solution, and will lead to significant progress in our understanding of the delicate balance of the fundamental forces that keep the biological machinery working, both for the emergence of the first replicase in the prebiotic soup and for the function of DNA in modern cells.

Prior to the studies on nucleobase complementarity, we studied the adsorption of the individual bases separately. Figure 1 shows STM images of the molecular networks formed by depositing molecules of G (Figure 1A and D), A (Figure 1B and E), and C (Figure 1C and F) on an Au(111) surface held at room temperature. The STM images were recorded at 100–150 K to prevent undesired diffusion events and to enhance the stability of the tunneling junction. The corrugation of the potential energy landscape for nucleobases adsorbed on Au(111) is sufficiently small that the molecules could easily diffuse and form the observed self-assembled molecular networks, which arise from hydrogen bonding between the peripheral functional groups of the nucleobase molecules, consistent with previous studies of G,[8] C,[9] T,[10] and A networks[11] on solid substrates.[12–15] Nucleobase molecules adsorbed on noble-metal (111) surfaces are found

[*] Dr. R. Otero,[+] [$] Dr. W. Xu,[$] Dr. M. Lukas,[++] [$] Prof. E. Lægsgaard, Prof. I. Stensgaard, Prof. Dr. F. Besenbacher
Interdisciplinary Nanoscience Center (iNANO), Center for DNA Nanotechnology (CDNA), and Department of Physics and Astronomy, University of Aarhus, Ny Munkegade, 8000 Aarhus C (Denmark)
Fax: (+45) 8612 0740
E-mail: fbe@inano.dk
Dr. R. E. A. Kelly,[+++] [$] Dr. L. N. Kantorovich
Department of Physics, School of Physical Sciences and Engineering, King's College London (UK)
Prof. J. Kjems
Interdisciplinary Nanoscience Center (iNANO), Center for DNA Nanotechnology (CDNA), and Department of Molecular Biology, University of Aarhus (Denmark)
[+] Present Address: Dep. De Física de la Materia Condensada, Universidad Autónoma de Madrid (Spain)
[++] Present Address: Institute for Nanotechnology, Forschungszentrum Karlsruhe (Germany)
[+++] Present Address: Department of Physics and Astronomy, University College London (UK)
[$] These authors contributed equally to this work.

[**] We acknowledge the financial support from the Danish Ministry for Science, Technology and Innovation through the iNANO Center, from the Danish Research Councils, and from the Carlsberg Foundation. R.O. and M.L. acknowledge support through EU individual Marie Curie Fellowship. Computational resources were obtained from the Danish Center for Scientific Computing. R.E.A.K. is grateful to the EPSRC for financial support, Grant No. GR/P01427/01. Finally, we acknowledge the financial support through EU-RTN AMMIST.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200803333.

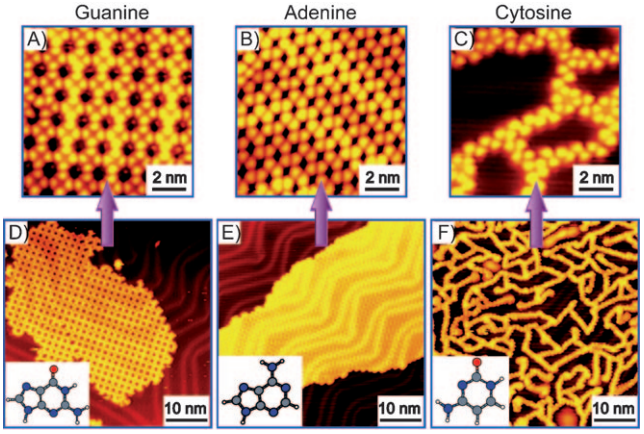


***Figure 1.*** STM images of single DNA bases deposited on Au(111) at room temperature. Guanine (A, D) and adenine (B, E) show 2D island growth. The herringbone reconstruction, known for clean Au(111), can clearly be seen as modulation on the top of the purine structures, that is, corrugation of the surface potential is small. Cytosine (C, F) grows into 1D filaments consisting of zigzag branches and ringlike structures. Blurs in areas confined by branches are due to mobile C molecules. The insets show the chemical structure of the molecules with O, N, C, and H atoms as red, blue, gray, and small white circles. All images were acquired in constant-current mode ($I_t = -0.3$–$0.7$ nA, $V_t = -800$–$1500$ mV).

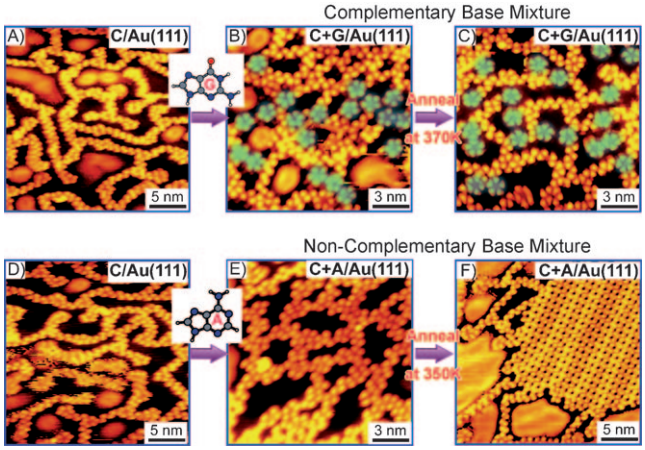


***Figure 2.*** Co-deposition experiment for complementary C + G (A–C) and noncomplementary C + A (D–F) bases. During the first step, similar amounts of C were deposited at room temperature in each case (A, D). After deposition of G (B) a sharp increase in the number of fivefold rings is found (indicated by the green shading), which was not observed after deposition of A (E). C + G and C + A mix on co-deposition (B, E). After heating, the complementary C + G mixture remains disordered (C), while the noncomplementary C + A mixture segregates into A islands and C zigzag branches (F).

to adopt a flat-lying adsorption geometry in which the aromatic rings of the bases are parallel to the substrate surface.[8–12,16–18] The hydrogen bonding is relatively unaffected by the interaction with the Au(111) surface.[11]

For the G and A molecules the high-resolution STM images clearly show the formation of self-assembled two-dimensional (2D) well-ordered nanostructures (Figure 1 A and B). The G molecules self-assemble into a hydrogen-bonded G-quartet network similar to those found in G-quadruplex DNA,[8] whereas the A molecules self-assembled into a hexagonal network. In contrast, cytosine molecules do not grow into regular 2D islands (Figure 1 C). At low coverages individual C molecules and small C clusters are very mobile and cannot be imaged by STM even at low temperatures (ca. 100 K). However, when the C coverage is increased, the C molecules form a disordered molecular network of 1D zigzag filaments interconnected by five- and sixfold rings, which appears to have glasslike characteristics.[9]

Next, co-deposition experiments on the binary nucleobase mixtures were performed. First we deposited C molecules at room temperature, which resulted in the typical zigzag branches and ringlike structures (Figure 2 A and D), as described above, and from the STM images the C coverage was determined. Subsequently, either the complementary base (G) or the noncomplementary base (A) was deposited onto the partially C covered surface (held at room temperature), and afterwards the surfaces were imaged by STM at low temperatures (Figure 2 B and E). Finally, the samples were heated for 10–15 min at a temperature below the lower desorption temperature of the two bases contained in the respective mixture, and the surfaces were imaged by STM at low temperatures again (Figure 2 C and F). STM images of such a sequential co-deposition and heating sequence are shown in Figure 2 A–C for C + G and Figure 2 D–F for C + A.

In both sets of experiments we deposited similar amounts of C in the first step, and after deposition of the purine molecules (G or A), the structures observed in the STM images for both binary mixtures seem to be very similar (Figure 2 B and E). However, a detailed and thorough analysis revealed that upon G deposition a large number of fivefold rings are observed (Figure 2 B, indicated in green) as compared to the case when A was deposited (Figure 2 E). Notably, both purine molecules (G and A) are incorporated into the pre-existing C network (see detailed analysis below), and thus both G–C and A–C pairs must exist.

Surprisingly, dramatic differences are observed after heating the binary mixtures. When the complementary C + G mixture is heated to 373 K for 10 min, the structure appears similar to that observed prior to heating (Figure 2 C), and we can thus conclude that the binary mixture of C + G does not change its characteristics upon heating at 373 K. However, when the sample is heated to even higher temperature (400 K), the C molecules are desorbed, and only the pure G molecules are left on the surface (Figure S1, Supporting Information). This is in agreement with the observation by Demers et al.[19] that C molecules are desorbed at a lower temperature than G molecules on Au(111). On the contrary, in the case of the noncomplementary C + A mixture, after heating to 353 K for 10 min, we mainly find large islands and zigzag branches, which appear to be segregations of the self-assembled structure of pure A and pure C (Figure 2 F and Figure S2 in Supporting Information).

Such experiments were carried out with varying amounts of C, G, and A, and they gave the same qualitative results, and we can thus conclude that a gentle thermal treatment

  

"induces" C to selectively bind to its complementary partner G rather than to A. From the observed dissociation temperatures for C + G and C + A networks of 400 K and 350 K, respectively, we can estimate that the CG pair is at least 15% more stable than the CA pair.

In the next step a thorough analysis of the high-resolution STM images was performed to obtain further atomic-scale insight into the enhanced stability of the complementary C + G mixtures. From the STM results of the C + G mixture depicted in Figure 3A, we can clearly distinguish two different kinds of molecules: some appear as larger, elongated triangular protrusions, assigned previously to G molecules,[8] and others as small equilateral triangles. The apparent larger triangular molecules were found to be incorporated in the filament structures, basically forming a local dimer structure with one of the small molecules. Furthermore, as depicted in Figure 3A, these dimers are also commonly found embedded in the fivefold rings, many of which consist of one larger triangular molecule and four smaller molecules closing the fivefold ring. We therefore conclude that most of the fivefold rings of the C + G mixture are composed of one G molecule, with an orientation that can be deduced from the shape of the triangle, and four C molecules. A similar analysis reveals that the G–C pairs exist in the filament structures as well (e.g., see Figure 3A), that is, G and C molecules are intermixed in the final binary mixture.

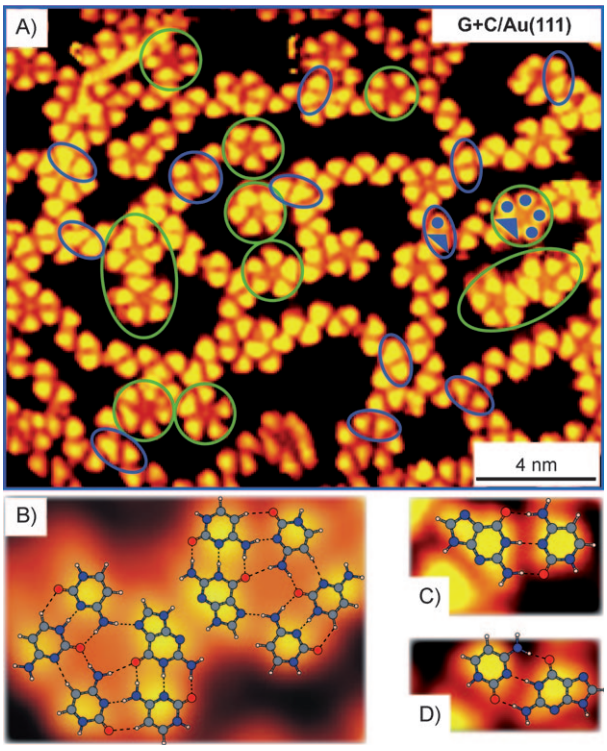


***Figure 3.*** Detailed analysis of the complementary mixture and comparison with DFT calculations. A) High-resolution STM image of the complementary C + G structure after heating. The blue ovals indicate G–C pairs involved in filamentary structures, and the green circles and ovals fivefold rings containing G–C pairs. The blue triangles indicate G, and the blue balls C molecules. B) Enlarged image (2.2 × 3.2 nm) showing two fivefold rings, each of which contains one WC G–C base pair. The superimposed structures are the relaxed minimum-energy DFT calculated geometries, which fit quite well to the experimental STM image. C, D) The enlarged STM images show two G–C pairs with different chiralities, which are superimposed by the calculated WC G–C base-pair structure.

To obtain further fundamental insight into the atomic-scale structures of the C + G mixture, DFT calculations were performed on several models, all selected with the criterion of maximizing the number of hydrogen bonds in the structures. The starting point in the analysis is the Watson–Crick (WC) G–C base pair (with a stabilization energy of −1.21 eV[20–23]), which has the highest stability amongst all nucleobases.[20,21] The pairs next in stability, G–G (−1.12 eV) and G–C (−1.06 eV), bind to each other through analogous hydrogen-bonding groups. However, being less stable, they would be less probable in the mixture as well. All other G–C, C–C, and G–G pairs are considerably less favorable and will only be realized to make use of the binding sites exposed around the more stable pairs. Therefore, most of the G–C pairs in the mixture should be WC base pairs. Using this WC base pair as the main building block, we constructed molecular models for the fivefold ring containing a single G and four C molecules. In Figure 3B–D the DFT relaxed structures of the ring and the WC base pair are superimposed on the recorded STM images of the C + G mixture, and a very good fit is found. Most of the observed rings and filament structures thus contain the WC G–C base pairs, and the fact that the WC base pairs are the smallest stable structures that can be formed in a C + G mixture explains the observed enhanced stability of C + G mixtures to heating. A possible scenario is that only the WC G–C pairs survive heating, while the other, much weaker C–C, G–C, and G–G pairs break up, so that the molecules move freely on the surface. When the system is cooled down again, a newly established mixture phase is formed, triggered by the already existing WC G–C pairs, which serve as precursors of the new structure.

Similar STM experiments with complementary A and T (instead of C and G) were not successful, as it was difficult to distinguish the molecules in the mixture. In fact, A + T complementarity has only been conclusively observed between amphiphile nucleotides on graphite.[24]

In the case of the noncomplementary mixture C + A, the binding energies for the most stable C–A pairs, between −0.75 and −0.88 eV, are slightly smaller than or comparable to those of the most stable C–C (−0.99 eV)[25] and A–A (−0.86 eV) pairs.[26] In this case all pairs break up on heating, and the whole network is completely disintegrated. Since A molecules can form highly stable hexagonal 2D islands with more highly coordinated molecules than in the initial filamentary structure of the mixture, these islands form preferably. The C molecules coexisting on the surface have no choice but to form C filaments and attach to the boundaries of the A islands. Comparing the most stable G–C and A–C pairs, we find that the G–C pair is more stable than the A–C pair by about 38%, which is larger than the experimentally imposed lower bound mentioned above.

In conclusion, our combined STM and DFT studies revealed that the resilience of the complementary nucleobase mixture to heating is due to the formation of stable WC base pairs between adsorbed G and C molecules on a flat solid

surface. Since this molecular recognition process is observed under extreme UHV conditions in the absence of water and phosphate backbone, it is most likely driven solely by the energy difference between the hydrogen-bonded A–C and G–C pairs. The observed formation of specific WC base pairs on a solid surface shows that the DNA backbone may not necessarily be a prerequisite for specific WC base pairing, as a flat surface, which effectively serves as a reservoir for a 2D gas of base molecules, may also trigger recognition between complementary bases. This finding may have important implications for creating models explaining the origin of life. First, it suggests that some planar solid surfaces, which may have been present in the prehistorical soup, could indeed have played an active role in catalyzing the formation of normal WC base pairs with reasonable accuracy. This type of recognition requires the ability of the surface to strongly restrict molecules from leaving it and, at the same time, to provide for their free mobility across the surface and thus give them greater opportunities to arrange themselves correctly for planar base pairing to take place. Second, it is reasonable to imagine that the heterocyclic ring structures of WC base pairs, like those observed in this study, may have stacked on each other, even in the absence of the sugar ring and backbone structures. It is possible that such short stacks of base pairs may act as a primitive catalyst for the synthesis of a covalent backbone, which presumably was a prerequisite for the emergence of the first primitive form of replication. This hypothesis would help to explain one of the paradoxes of life: How could a pool of chemically much more complicated and stereospecific nucleotides have emerged before the appearance of the first replicase.

## Experimental Section

Samples were prepared and investigated under UHV conditions to ensure a minimum of disturbing impurities. The bases were purchased as powders from Sigma Aldrich (purities: G 98%, A 99%, C 97%) and were degassed for several hours in UHV prior to deposition. The Au(111) single crystal was cleaned by cycles of 1.5-keV $Ar^+$ sputtering and consecutive heating to 820 K until a clean surface, indicated by the well-known herringbone reconstruction, was obtained. Bases were deposited by sublimation onto the clean surface held at room temperature. STM investigations were carried out in a home-built liquid-nitrogen-cooled Aarhus STM[27] at temperatures ranging from 100 to 300 K. The STM experiments were carried out at low temperatures (100–150 K) to minimize the surface mobility of the molecules. Since no definite chemical and orientational assignment is possible from the STM images, DFT-based calculations were performed by the SIESTA method.[28] The theoretical calculations do not include the Au(111) surface, since the interaction between the surface and molecules has been found to have a negligible corrugation across the surface.[8,11]

Received: July 9, 2008
Revised: August 24, 2008
Published online: November 10, 2008

**Keywords:** density functional calculations · molecular recognition · nucleobases · scanning probe microscopy · surface chemistry

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.
[2] J. L. Bada, A. Lazcano, *Science* **2002**, *296*, 1982.
[3] S. J. Sowerby, W. M. Heckl, *Origins Life Evol. Biosphere* **1998**, *28*, 283.
[4] L. E. Orgel, *Nature* **1992**, *358*, 203.
[5] G. F. Joyce, *Nature* **1989**, *338*, 217.
[6] a) J. V. Barth, *Annu. Rev. Phys. Chem.* **2007**, *58*, 375; b) J. V. Barth, G. Costantini, K. Kern, *Nature* **2005**, *437*, 671.
[7] S. De Feyter, C. F. De Schryver, *Chem. Soc. Rev.* **2003**, *32*, 139.
[8] R. Otero, M. Schöck, L. M. Molina, E. Lægsgaard, I. Stensgaard, B. Hammer, F. Besenbacher, *Angew. Chem.* **2005**, *117*, 2310; *Angew. Chem. Int. Ed.* **2005**, *44*, 2270.
[9] R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Science* **2008**, *319*, 312.
[10] W. Xu, R. E. A. Kelly, R. Otero, M. Schöck, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2007**, *3*, 2011.
[11] R. E. A. Kelly, W. Xu, M. Lukas, R. Otero, M. Mura, Y.-J. Lee, E. Lægsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2008**, *4*, 1494.
[12] a) M. Furukawa, H. Tanaka, T. Kawai, *J. Chem. Phys.* **2001**, *115*, 3419; b) T. Kawai, H. Tanaka, T. Nakagawa, *Surf. Sci.* **1997**, *386*, 124; c) M. Furukawa, H. Tanaka, T. Kawai, *Surf. Sci.* **1997**, *392*, L33; d) M. Furukawa, H. Tanaka, T. Kawai, *Surf. Sci.* **2000**, *445*, 1.
[13] L. M. A. Perdigão, P. A. Staniec, N. R. Champness, R. E. A. Kelly, L. N. Kantorovich, P. H. Beton, *Phys. Rev. B* **2006**, *73*, 195423.
[14] a) M. Edelwirth, J. Freund, S. J. Sowerby, W. M. Heckl, *Surf. Sci.* **1998**, *417*, 201; b) S. J. Sowerby, M. Edelwirth, W. M. Heckl, *J. Phys. Chem. B* **1998**, *102*, 5914; c) W. M. Heckl, D. P. E. Smith, G. Binnig, H. Klagges, T. W. Hansch, J. Maddocks, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 8003; d) S. J. Sowerby, C. A. Cohn, W. M. Heckl, N. G. Holm, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 820.
[15] Q. Chen, N. V. Richardson, *Nat. Mater.* **2003**, *2*, 324.
[16] N. J. Tao, J. A. DeRose, S. M. Lindsay, *J. Phys. Chem.* **1993**, *97*, 910.
[17] S. Piana, A. Bilic, *J. Phys. Chem. B* **2006**, *110*, 23467.
[18] K. F. Domke, D. Zhang, B. Pettinger, *J. Am. Chem. Soc.* **2007**, *129*, 6708.
[19] L. M. Demers, M. Östblom, H. Zhang, N.-H. Jang, B. Liedberg, C. A. Mirkin, *J. Am. Chem. Soc.* **2002**, *124*, 11248.
[20] R. E. A. Kelly, L. N. Kantorovich, *J. Phys. Chem. C* **2007**, *111*, 3883.
[21] J. Sponer, P. Jurecka, P. Hobza, *J. Am. Chem. Soc.* **2004**, *126*, 10142.
[22] C. F. Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *Chem. Eur. J.* **1999**, *5*, 3581.
[23] S. Xu, M. Dong, E. Rauls, R. Otero, T. R. Linderoth, F. Besenbacher, *Nano Lett.* **2006**, *6*, 1434.
[24] I. Bestel, N. Campins, A. Marchenko, D. Fichou, M. W. Grinstaff, P. Barthélémy, *J. Colloid Interface Sci.* **2008**, *323*, 435.
[25] R. E. A. Kelly, Y. J. Lee, L. N. Kantorovich, *J. Phys. Chem. B* **2005**, *109*, 22045.
[26] R. E. A. Kelly, Y. J. Lee, L. N. Kantorovich, *J. Phys. Chem. B* **2005**, *109*, 11933.
[27] E. Laegsgaard, L. Osterlund, P. Thostrup, P. B. Rasmussen, I. Stensgaard, F. Besenbacher, *Rev. Sci. Instrum.* **2001**, *72*, 3537.
[28] a) P. Ordejon, E. Artacho, J. M. Soler, *Phys. Rev. B* **1996**, *53*, 10441; b) J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, D. Sanchez-Portal, *J. Phys. Condens. Matter* **2002**, *14*, 2745.

  