J|A|C|S ARTICLES

Published on Web 09/15/2010

Clay Minerals Mediate Folding and Regioselective Interactions of RNA: A Large-Scale Atomistic Simulation Study

Jacob B. Swadling, Peter V. Coveney,* and H. Christopher Greenwell†

Centre for Computational Science, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom

Received May 13, 2010; E-mail: p.v.coveney@ucl.ac.uk

Abstract: Since a mineral-mediated origin of life was first hypothesized over 60 years ago, clays have played a significant role in origins of life studies. Such studies have hitherto rarely used computer simulation to understand the possible chemical pathways to the formation of biomolecules. We use molecular dynamics techniques, performed on supercomputing grids, to carry out large-scale simulations of various 25-mer sequences of ribonucleic acid (RNA), in bulk water and with aqueous montmorillonite clay over many tens of nanoseconds. Hitherto, there has only been limited experimental data reported for these systems. Our simulations are found to be in agreement with various experimental observations pertaining to the relative adsorption of RNA on montmorillonite in the presence of charge balancing cations. Over time scales of only a few nanoseconds, specific RNA sequences fold to characteristic secondary structural motifs, which do not form in the corresponding bulk water simulations. Our simulations also show that, in aqueous Ca^{2+} environments, RNA can tether to the clay surface through a nucleotide base, leaving the 3′-end of the strand exposed, providing a mechanism for the regiospecific adsorption and elongation of RNA oligomers on clay surfaces.

1. Introduction

Montmorillonite clay is a principal alteration product of volcanic ash, and it is relatively widespread within bentonite deposits in the present day Earth's crust. It is postulated that similar alteration processes would also have occurred on the early Earth,[1] with interpetation of the oxygen isotope composition of Hadean zircons indicating that the conditions needed for the formation of sodium montmorillonite existed at up to 4.4 billion years ago.[2,3] Under laboratory conditions, montmorillonite has been shown to catalyze the formation of long ribonucleic acid (RNA) oligomers, along with a large number of other organic reactions.[4] Montmorillonite is a 2:1 clay consisting of an octahedral alumina (AlO_6) sheet sandwiched between two tetrahedral silica (SiO_4) sheets. Naturally occurring montmorillonite, such as Wyoming montmorillonite, contains partial isomorphic substitutions in both tetrahedral and octahedral sheets, which confer a net negative charge. Silicon (4+) ions in the tetrahedral sheets are commonly substituted with aluminum (3+) ions, and octahedrally coordinated aluminum (3+) by magnesium (2+) and iron (2+) ions. The negative charge is balanced by various exchangeable cations, often sodium (Na^+) and calcium (Ca^{2+}) ions, which reside on the surfaces and within the interlayer region of the clay. Work performed by Joshi et al. has shown that of 22 chemically distinct varieties of natural montmorillonite, the Wyoming type has greatest catalytic activity.[5]

The ability of montmorillonite to catalyze the formation of RNA oligomers has been demonstrated in a number of experiments by Huang et al.[6] In their laboratory studies Ferris and co-workers used artificial activated phosphorimidazolides of nucleosides together with homoionic Na^+ montmorillonite to form oligomers of 6−14 mers, in which montmorillonite was shown to enhance the rate constant for oligomer formation by 100−1000 times over the hydrolysis of the imidazole activating group.[7,8] Ferris et al. then experimented with the use of a primer strand of 10 mers which produced 30−40 mers.[9] By changing the activating group from imidazole to 1-methyladenine, Huang et al. formed single strands of up to 50 mers of polyadenine and polyuracil in a couple of days, without the need for a primer strand. RNA sequences of this length are sufficiently long to exhibit fidelity in replication as well as catalytic activity.[6] It is important to note that the chemistry reported could not have transpired on the primitive Earth because it involves activating chemical bonds in the manner of modern organic chemistry. How such polymerization could have occurred in the prebiotic era remains to be fully understood.

As each montmorillonite clay sheet carries a net negative charge and each RNA strand has negatively charged phosphate groups, even the basic mechanism by which the RNA interacts

† Present address: Department of Chemistry, Durham University South Road, Durham, DH1 3LE, United Kingdom.

(1) Zaikowski, L.; Friedrich, J. *Chemical Evolution across Space and Time From the Big Bang to Prebiotic Chemistry*; OUP: Oxford, UK, 2008.
(2) Watson, E. B.; Harrison, T. M. *Science* **2005**, *308*, 841.
(3) Wilde, S.; Valley, J.; Peck, W.; Graham, C. *Nature* **2001**, *409*, 175–178.
(4) Nikalje, M. D.; Phukan, P.; Sudalai, A. *Org. Prep. Proced. Int.* **2000**, *32*, 1.
(5) Joshi, P. C.; Aldersley, M. F.; Delano, J. W.; Ferris, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 13369–13374.
(6) Huang, W. H.; Ferris, J. P. *J. Am. Chem. Soc.* **2006**, *128*, 8914–8919.
(7) Ferris, J. P.; Ertem, G. *J. Am. Chem. Soc.* **1993**, *115*, 12270–12275.
(8) Kawamura, K.; Ferris, J. P. *Origins Life Evol. Biosphere* **1999**, *29*, 563–591.
(9) Ferris, J. P.; Hill, A. R.; Liu, R. H.; Orgel, L. E. *Nature* **1996**, *381*, 59–61.

10.1021/ja104106y

with clays needs clarification. Franchi et al. have proposed that cations play a key role in the adsorption of the strands on clay surfaces.[10] The cations could potentially mediate the interaction by screening the negative charge using monovalent cations, or by creating a divalent cation bridge between the two layers of negative charge from the nucleic acid and mineral surface. These authors have demonstrated that the quantity of single-stranded adenine and uracil RNA (poly[A] and poly[U]) adsorbed is greater in the presence of Ca^{2+} than Na^{+} ions.

One of the leading theories concerning the origin of life is the RNA world hypothesis, based on the assumption that an RNA-based world preceded the world we live in now, in which according to the central dogma of modern biology DNA makes RNA which in turn makes proteins.[11] The hypothesis is attractive in part because it only requires the formation of one type of biopolymer. The RNA world hypothesis has gained significant support from the work of Cech and Altman, who won the 1989 Nobel prize in chemistry for the discovery of a group of catalytic RNA molecules called ribozymes.[12] Steitz, along with many others, elucidated the structure of the ribosome, which is the molecular machine responsible for making proteins and amino acids,[13,14] showing that the ribosome is a ribozyme. More recent support has come from Powner et al. who synthesized purine-based RNA.[15] Montmorillonite catalysis could potentially resolve some of the problems in RNA synthesis by concentrating and polymerizing the RNA and producing stereoenantiomers.[16] Much experimental work has been performed on the interaction of nucleic acids (both single- and double-stranded RNA and DNA) interacting with mica surfaces.[17,18] Although the mica/nucleic acid system is not identical to that of montmorillonite/nucleotide systems, the mica surface can be thought of as an analogous system as both mica and montmorillonite have negatively charged surfaces. For many years montmorillonite has been associated in part with the RNA world hypothesis, and therefore the origins of life on Earth. Indeed, the idea that clay minerals played an important role in the early chemical processes leading up to the origin of life was proposed independently by Goldschmidt[19] and Bernal[20] as long ago as 1947.

Even today, however, the structure and conformation of RNA interacting with montmorillonite remains uncertain, as well as the mechanism of elongation of the polymer by addition of nucleic acid monomers. Research into origins of life has hitherto rarely used simulation techniques to understand the possible chemical pathways to the formation of biomolecules.[21−23] However, over the past 10−15 years, molecular simulation has played an increasingly important role in clay chemistry,[24,25] as well as in the study of RNA.[26] Computer simulations complement current experimental techniques by providing atomic-scale resolution of such systems. In particular, simulation provides insight into how molecules adsorb on and intercalate within clays.

In this paper we explore the mechanism by which RNA adsorbs on an aqueous Wyoming montmorillonite mineral surface, using fully atomistic large-scale classical molecular dynamics techiques. We look at the interaction of various RNA molecules, of differing base sequence, with the mineral surface in the presence of Ca^{2+} or Na^{+} charge balancing cations. We investigate how the differing counterions alter the RNA conformation, folding kinetics, and structure, and compare its behavior on montmorillonite to that in aqueous solution. We describe how RNA changes its structure and conformation when interacting with a montmorillonite surface and contrast the structure and conformation in bulk water.

In section 2 of this paper we describe techniques that we have employed to simulate the clay/nucleic acid systems reported in this paper. Section 3 discusses the results obtained in these simulations, while section 4 draws conclusions from this study.

2. Methods

This section discusses the techniques that we have employed to simulate the aqueous nucleic acids and clay systems in this paper, including the choice of potentials needed to describe the interaction between RNA and montmorillonite. We have used fully atomistic, large-scale molecular dynamics techniques to simulate these flexible systems containing hundreds of thousands of atoms. A collection of three federated supercomputing grids coupled to local resources at UCL were utilized in order to perform these simulations, and their associated data analysis and transfer.

2.1. Model Construction. The montmorillonite used in this study was derived from an initial structure with the chemical formula $Na_3[Al_{14}Mg_2][Si_{31}Al]O_{80}(OH)_{16}$ shown in Figure 1a, which is consistent with naturally occurring Wyoming montmorillonite. Partial substitutions were created in both tetrahedral and octahedral sheets, although there is no evidence that the negative charge arising from partial substitution of Al^{3+} ions for Mg^{2+} ions in the octahedral sheet is implicated in the bonding between RNA and montmorillonite. The montmorillonite structure was constructed using the Materials Studio (Accelrys) software.[27] The montmorillonite structure was chosen as it is widely used in experiments[10,28] and it has been shown to possess catalytic activity with respect to RNA oligomerization.[5]

The sizes and sequences of the RNA molecules used in this study are given in Table 1 and shown in Figure 1b. Table 1 uses standard notation for displaying RNA sequences, where the strand is displayed in the 5′−3′ direction exhibiting individual nucleotide bases as A, U, C, and G for adenine, uracil, cytosine, and guanine, respectively. All the polymers in models **I**−**XII** are 25 base groups in length and are single-stranded. All the RNA molecules, denoted **A**, **B**, **C**, and **dA**, were assembled using the Nucleic Acid Builder

(10) Franchi, M.; Ferris, J. P.; Gallori, E. *Origins Life Evol. Biosphere* **2003**, *33*, 1–16.
(11) Dworkin, J. P.; Lazcano, A.; Miller, S. L. *J. Theor. Biol.* **2003**, *222*, 127–134.
(12) Cech, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4360–4363.
(13) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905–920.
(14) Cech, T. *Science* **2000**, *289*, 878–879.
(15) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *459*, 239–242.
(16) Joshi, P. C.; Pitsch, S.; Ferris, J. P. *Origins Life Evol. Biosphere* **2007**, *37*, 3–26.
(17) Pastre, D.; Pietrement, O.; Fusil, P.; Landousy, F.; Jeusset, J.; David, M. O.; Hamon, C.; Le Cam, E.; Zozime, A. *Biophys. J.* **2003**, *85*, 2507–2518.
(18) Hansma, H. G.; Revenko, I.; Kim, K.; Laney, D. E. *Nucleic Acids Res.* **1996**, *24*, 713–720.
(19) Goldschmidt, V. *New Biol.* **1952**, *12*, 97–105.
(20) Bernal, J. D. *Proc. Phys. Soc. London Sect.* **1949**, *62*, 537–558.
(21) Thyveetil, M.-A.; Coveney, P. V.; Greenwell, H. C.; Suter, J. L. *J. Am. Chem. Soc.* **2008**, *130*, 4742–4756.
(22) Schuster, P. *Biol. Chem.* **2001**, *382*, 1301–1314.
(23) Mathew, D.; Luthey-Schulten, Z. *Origins Life Evol. Biosphere* **2010**, *40*, 303–317.
(24) Suter, J. L.; Coveney, P. V.; Greenwell, H. C.; Thyveetil, M.-A. *J. Phys. Chem. C* **2007**, *111*, 8248–8259.
(25) Thyveetil, M.-A.; Coveney, P. V.; Greenwell, H. C.; Suter, J. L. *J. Am. Chem. Soc.* **2008**, *130*, 12485–12495.
(26) Sponer, J.; Lankas, F. E. *Computational studies of RNA and DNA*; Springer: New York, 2006.
(27) Materials Studio, http://accelrys.com/products/materials-studio/.
(28) Ferris, J. P. *Origins Life Evol. Biosphere* **2002**, *32*, 311–332.

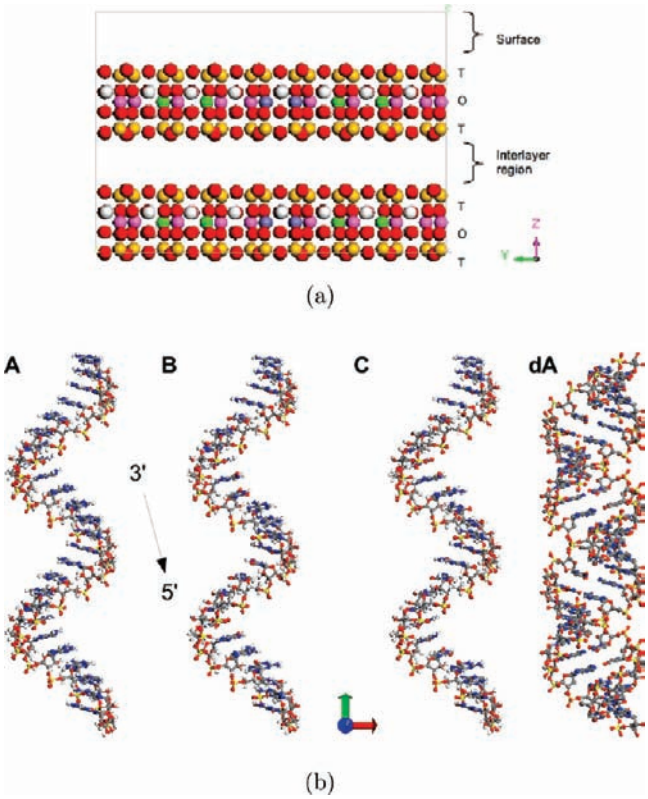

Figure 1. (a) Initial structure of the unit cell of Wyoming type montmorillonite with charge balancing ions and water removed for clarity. Atoms are colored as follows: Si (gold), O (red), H (white), Al (purple), Mg (green), and Fe (blue). The visualization depicts the tetrahedral and octahedral layers of the clay, as well as interlayer regions and the exposed surface. (b) RNA sequences **A**, **B**, and **C** show the three different 25-mer single strands of RNA, and **dA** depicts the double-stranded RNA used in this study. RNA sequence **A** is used in models **I**−**IV** and consists of 25 adenine base groups. RNA sequence **B** is used in models **V**−**VIII** and **C** in models **IX**−**XII**; both consist of different nucleotide base sequences described in more detail in Table 1. **dA** consists of a 25-mer adenine sequence and complementary uracil strand. Color scheme is as follows: O (red), H (white), N (blue), C (gray), and P (yellow). The visualization shows the RNA strands in the 3′ → 5′ direction, where the prime indicates the carbon atom on the ribose sugar component of the RNA monomer.

(NAB) which is part of AmberTools 1.2.[29] NAB produces nucleic acid structures in a hierarchical fashion on the basis of three successively applied steps: (i) transformation of nucleotide bases to achieve desired helical and base-pairing configurations, (ii) geometry optimization which allows molecular structures to be built that satisfy sets of distance constraints, and (iii) potential energy minimization. Models **I** and **II** consist of a single-stranded polyadenine RNA molecule, sequence **A**, aligned with the aqueous montmorillonite surface with differing cations neutralizing the negative charge from the partial substitutions in the clay and charged unprotonated phosphate groups. From the partial substitutions within the clay sheets, as shown in the clay chemical formula $Na_3[Al_{14}Mg_2][Si_{31}Al]O_{80}(OH)_{16}$, a net negative charge of 408− was generated on the periodic clay supercell structure. The RNA strands used in this study all bear a net negative charge of 24−. As such, a corresponding number of Na^+ or Ca^{2+} ions were added in order to neutralize the negative charge in all the systems. The phosphate groups were unprotonated and carry a negative charge as we are assuming a pH of 7 where all phosphate groups are known to be ionized. This sequence was chosen to mimic the strand used in the experiments performed by Franchi et al.[10] Models **III** and **IV** are the corresponding RNA in bulk aqueous solution (i.e., in the absence of montmorillonite); these are used to infer the effects of the clay surface on the structure, dynamics and conformation of the RNA molecule. In models **V**−**XII** two different single-stranded RNA sequences were used (**B** and **C**). The base sequences were taken from two stem−loop secondary structural motifs in the type III hammerhead ribozyme[30] (see Figure 2). The stem−loop motifs were chosen to observe the effects of the clay surface on the formation and stabilization of secondary structures. All sequences were also simulated using different cations and in bulk aqueous solution. The two cations used were monovalent Na^+ ions and divalent Ca^{2+} ions, in order to investigate the effects of cation charge density on RNA conformation and on mediating surface interactions; they were selected to mimic those used in experiments.[10] Model **Ia** was constructed in order to test the effect of the initial cation distribution on the simulation; in this model the cations were distributed randomly around sequence **A** using a stochastic algorithm. Cations in other models were entered manually into random positions within the simulation cell. Model **XIII** was simulated in order to provide a comparison with the double stranded DNA and RNA experiments preformed on a mica surface.

Distances of 20 Å in the clay models and 50 Å in the aqueous models were used to minimize the interaction between RNA and its periodic images. Over 25,000 water molecules were placed above the surface of the montmorillonite clay within the periodic simulation cell, giving it a length of 180 Å, sufficient to decouple the clay layer from its periodic image in the *z*-direction. The initial configuration of the RNA and the clay surface is shown in Figure 3. Charge-balancing cations were placed in the interlayer region of the clay and also between the RNA molecule and the surface manually. In the bulk water models cations were placed at random locations around the molecule using the 'solvate' and 'addions' routines within *Amber Tools*, which uses a stochastic algorithm.

2.2. Potential Parameterization. The RNA molecules were all parametrized with Amber's *Leap* module[32] and converted to LAMMPS input file format. The RNA and montmorillonite LAMMPS input files were aligned with the RNA molecule placed above the aqueous surface of the clay. The combination of the two input files is possible because both the Amber forcefield[33] used to model RNA and ClayFF forcefield[34] employed to describe montmorillonite have the same functional form regarding bonding parameters.

We use classical molecular dynamics simulations to study the interaction of RNA with montmorillonite. Interactions between minerals and organic molecules are difficult to describe with a single "standard" molecular mechanics forcefield. Most forcefields are parametrized for a specific organic or inorganic system. Previous simulations carried out on biomineral surfaces have coupled standard organic forcefields such as Amber[33] with inorganic forcefields[21,25] and also hybrid forcefields for minerals, for example ClayFF.[34] We use Lorentz−Berthelot mixing rules to supply the missing Lennard-Jones parameters when combining forcefields. These rules estimate the intermolecular parameters of the Lennard-Jones potential using an arithmetic average for the collision diameter and a geometric average for the well depth.[35] They work well in situations in which the dominant interactions are electrostatic, as is the case here. A cutoff of 10 Å was imposed on the Lennard-Jones interactions. Coulomb

(29) Case, D. A. et al., AMBER 10, http://ambermd.org/, 2008.
(30) Torres, R. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11077–11082.
(31) Gruber, A. R.; Lorenz, R.; Bernhart, S. H.; Neubock, R.; Hofacker, I. L. *Nucleic Acids Res.* **2008**, *36*, W70–W74.
(32) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
(33) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(34) Cygan, R. T.; Liang, J. J.; Kalinichev, A. G. *J. Phys. Chem. B* **2004**, *108*, 1255–1266.
(35) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press: New York, 1987.

Table 1. Simulation Cell Compositions and Dimensions for All the Systems Studied Herein[a]

model	RNA	sequence	cation	number of atoms	clay surface	starting cell dimensions (Å^3)	simulation time (ns)
I	**A**	(5′-AAAAAAAAAAAAAAAAAAAAAAAAA-3′)	Na^+	96.025	yes	82.74 × 143.28 × 180.62	33.866
Ia		(5′-AAAAAAAAAAAAAAAAAAAAAAAAA-3′)	Na^+	96.025	yes	82.74 × 143.28 × 180.62	23.116
II		(5′-AAAAAAAAAAAAAAAAAAAAAAAAA-3′)	Ca^{2+}	95.821	yes	82.74 × 143.28 × 180.62	34.008
III		(5′-AAAAAAAAAAAAAAAAAAAAAAAAA-3′)	Na^+	267.062	no	125.52 × 177.26 × 125.68	22.588
IV		(5′-AAAAAAAAAAAAAAAAAAAAAAAAA-3′)	Ca^{2+}	267.050	no	125.52 × 177.26 × 125.68	22.840
V	**B**	(5′-AAAGCAGAGUAAAGUCUCUGAGUAG-3′)	Na^+	96.011	yes	82.74 × 143.28 × 180.62	25.240
VI		(5′-AAAGCAGAGUAAAGUCUCUGAGUAG-3′)	Ca^{2+}	95.807	yes	82.74 × 143.28 × 180.62	31.340
VII		(5′-AAAGCAGAGUAAAGUCUCUGAGUAG-3′)	Na^+	261.084	no	125.52 × 177.26 × 125.68	20.786
VIII		(5′-AAAGCAGAGUAAAGUCUCUGAGUAG-3′)	Ca^{2+}	263.835	no	125.52 × 177.26 × 125.68	20.112
IX	**C**	(5′-AAAGUCUGGGCUAAGCCCACUGAUG-3′)	Na^+	96.011	yes	82.74 × 143.28 × 180.62	25.886
X		(5′-AAAGUCUGGGCUAAGCCCACUGAUG-3′)	Ca^{2+}	96.011	yes	82.74 × 143.28 × 180.62	31.484
XI		(5′-AAAGUCUGGGCUAAGCCCACUGAUG-3′)	Na^+	261.084	no	125.52 × 177.26 × 125.68	20.470
XII		(5′-AAAGUCUGGGCUAAGCCCACUGAUG-3′)	Ca^{2+}	261.084	no	125.52 × 177.26 × 125.68	20.866
XIII	**dA**	d(5′-AAAAAAAAAAAAAAAAAAAAAAAAA-3′)	Na^+	96.798	yes	82.74 × 143.28 × 180.62	24.792

[a] Three single-stranded RNA sequences were simulated, sequence **A** in models **I**−**IV**, sequence **B** in models **V**−**VIII**, and sequence **C** in models **IX**−**XII**. Each sequence was simulated in the presence of the montmorillonite surface and in aqueous solution, and simulated with Na^+ ions or Ca^{2+} ions as a countercation. Sequence **A** is a 25-mer polyadenine RNA. Sequences **B** and **C** are both 25-mer strands with mixed-base sequences taken from the hammerhead ribozyme which exhibits secondary structural features. A further model **XIII** was built consisting of a double strand of sequence **A**, denoted **dA**, on an aqueous montmorillonite surface. Model **Ia** is identical to model **I** except for the initial distribution of the charge balancing Na^+ ions on the surface of the clay, and was used to investigate the effects of changing the initial configuration of Na^+ ions on the ensuing structure and conformation of the RNA.

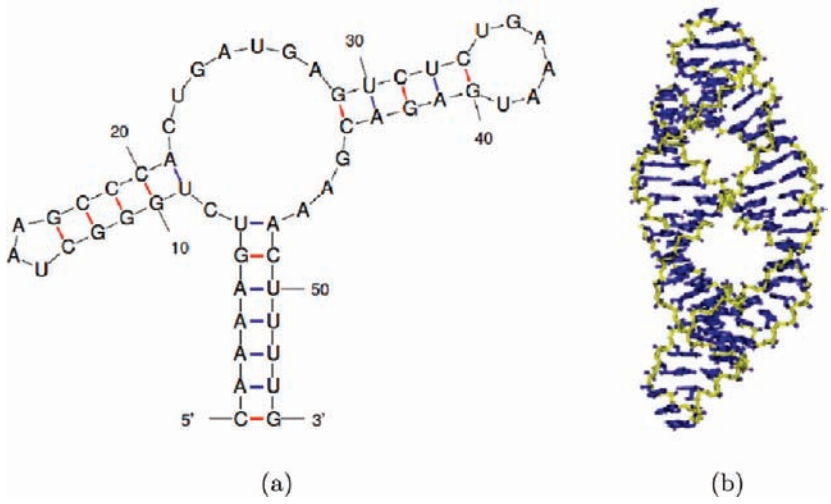

Figure 2. (a) Secondary structure of the hammerhead ribozyme (type III) and (b) the crystal structure of the RNA (type III) hammerhead ribozyme. Color scheme as in Figure 1. Secondary structure schematic in (a) was generated using the Vienna RNA software suite.[31] The tertiary structure of the ribozyme (b) was taken from the protein database file 1MME and visualized using VMD. Both (a) and (b) show the two stem−loop motifs which were used in sequences **B** and **C**; sequence **B** consists of the nucleotides 24−48, and sequence **C**, of nucleotides 3−27 shown in (a).

interactions were computed using the Ewald summation and the particle−particle, particle-mesh method with a precision value of 0.001 and a grid order of 4.[36] The Amber forcefield ff02[33] provided parameters for the partial charges and bonded interactions within the RNA molecule, while the ClayFF[34] forcefield furnished the parameters for atoms within the montmorillonite clay. The Amber forcefield has been used extensively to simulate RNA and reproduces well the structure and dynamics of nucleic acid moieties. MD simulations using the Amber forcefield have been performed on two ribosomal RNA loop motifs, accurately describing the Watson−Crick hydrogen-bonded base pairs.[37] Over a 14 ns time scale the Amber forcefield has been shown to provide structures consistent with NMR NOESY data.[38] Over longer time scales (∼100 ns), now becoming accessible through increasing computational power, the Amber forcefield may be somewhat less valid, and it is probably optimistic to claim the RNA parametrization to be fully "optimal". The limitations of nucleic acid potentials in MD are discussed in greater detail by Cheatham and Young.[39] ClayFF was developed to extend the CVFF[40] with parameters for clays and other layered minerals.[34] ClayFF produces good agreement with experiment for clays in terms of lattice parameters, water diffusion coefficients and far-infrared spectra and also with crystal parameters and vibrational spectra for a number of other layered minerals.[41] We have used ClayFF to simulate the structure and behavior of montmorillonite with a range of inorganic and organic species.[42] Both Amber and ClayFF use a harmonic potential for bond terms, while ClayFF is an ionic forcefield with no angle or dihedral terms, making the two forcefields relatively simple to combine; as noted above, we have successfully used them in the past to simulate

(36) Hockney, R. W.; Eastwood, J. R. *Computer simulation using particles*; Mc Graw Hill: 1981.
(37) Reblova, K.; Spackova, N.; Stefl, R.; Csaszar, K.; Koca, J.; Leontis, N. B.; Sponer, J. *Biophys. J.* **2003**, *84*, 3564–3582.
(38) Arthanari, H.; McConnell, K. J.; Berger, R.; Young, M. A.; Beveridge, D. L.; Bolton, P. H. *Biopolymers* **2003**, *68*, 3–15.
(39) Cheatham, T. E., III; Young, M. A. *Biopolymers* **2001**, *56*, 232–256.
(40) Dauber-Osguthorpe, P.; Roberts, V.; Osguthorpe, D.; Wolff, J.; Genest, M.; Hagler, A. *Proteins: Struct., Funct. Bioinf.* **1988**, *4*, 31–47.
(41) Cygan, R.; Greathouse, J.; Heinz, H.; Kalinichev, A. *J. Mater. Chem.* **2009**, *19*, 2470–2481.
(42) Suter, J.; Anderson, R.; Greenwell, H.; Coveney, P. *J. Mater. Chem.* **2009**, *19*, 2482–2493.

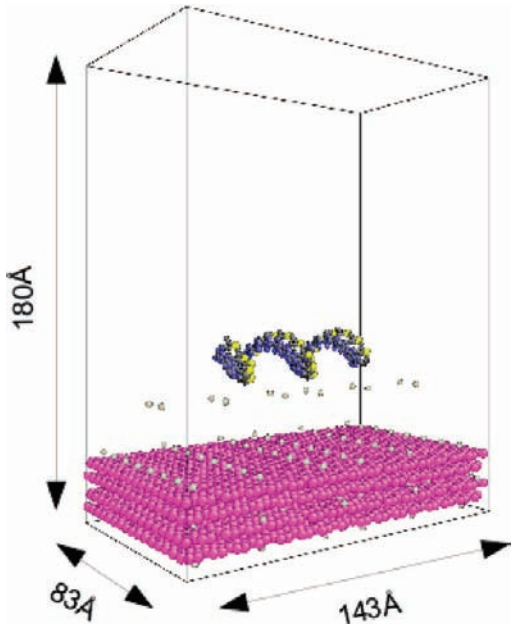

Figure 3. Starting structure of model **II** in Table 1, the silicon atoms within montmorillonite being shown as magenta spheres; remaining clay atoms have been hidden to aid viewing. Nitrogen, carbon, and phosphorus atoms belonging to the RNA molecule are colored the same as in Figure 1. Ca^{2+} ions are shown as gray spheres which were placed manually at random locations. Water molecules are not displayed. The helical structure of the RNA strand was created using the Nucleic Acid Builder, described in more detail in section 2.1.

DNA−clay systems.[21,25] Water molecules are described using the flexible single-point charge (SPC) model.[43] All models were simulated at ambient conditions of 300 K and 1 atm, the conditions for which the forcefields were parametrized.

2.3. Molecular Dynamics Techniques. The Large-scale Atomistic/Molecular Massively Parallel Simulator (LAMMPS)[44] was used to simulate all the models described here, due to its highly scalable nature. The models were energy-minimized using the steepest-descent method. A thermalization step was run, heating the system up from 50 K to 300 K over 0.01 ns. The whole system was allowed to evolve and equilibrate using the NPT ensemble. Equilibration was deemed to have been established by monitoring the potential energy of the system as a function of time, the radius of gyration of the RNA molecule, and by looking for the cell dimensions to take on stable average values; the plots showing the system reaching equilibrium have been included within the Supporting Information. Fourteen models were simulated in total. All simulations were run on the geographically distributed computational resources described in section 2.4. Clay models were run on 512 processing cores and the bulk water models on 1024 cores. The number of cores was selected on the basis of the number of atoms of each model and the scaling properties of the LAMMPS code. Each model was simulated in production conditions for the specified number of nanoseconds following equilibration, shown in Table 1. Models **I** and **II** were simulated for 30 ns, where it was shown that equilibrium had been established for over 20 ns; as a consequence, all other simulations were run for at least 20 ns.

The structure and arrangement of the various models were analyzed using methods reported in our previous studies,[21,25,45] including radial distribution functions (RDF) and atomic density profiles. A general measure of diffusive molecular motion is the mean-squared displacement (MSD), which was used to compute the diffusion of the RNA molecule. The MSD was calculated using:

$$\langle r^2(t)\rangle = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{r}_{i,0})^2 \quad (1)$$

where $\mathbf{r}_i$ is the current position of atom i, and $\mathbf{r}_{i,0}$ is its initial position. The diffusion was subsequently estimated from the diffusion coefficient using:

$$\langle r^2(t)\rangle = 6Dt + C \quad (2)$$

where C and D are constants; D is the diffusion coefficient and defines the rate of diffusion of the molecule, and t is time. The diffusion coefficient can be found from the gradient of the mean-squared displacement as a function of time. The radius of gyration, r_G, estimates the molecule's effective size during the simulation:

$$r_G = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \langle\mathbf{r}\rangle)^2} \quad (3)$$

where $\langle\mathbf{r}\rangle$ is the mean position of all N RNA atoms. Principal component analysis (PCA) was used to describe the main changes in RNA conformation with time. PCA is widely exploited to reduce the dimensionality of an MD trajectory, and it identifies the dominant collective modes of motion of molecules.[46] In a Cartesian coordinate system, the covariance matrix can be defined as:

$$\mathbf{C} = \langle(\mathbf{r} - \langle\mathbf{r}\rangle)(\mathbf{r} - \langle\mathbf{r}\rangle)^T\rangle \quad (4)$$

where $\mathbf{r}$ represents the atomic position of the RNA molecule in a $3N$ dimensional configuration space, $\langle\mathbf{r}\rangle$ is the mean position of atoms over all snapshots, and the superscript T denotes the matrix transpose. In PCA, the eigenvectors and corresponding eigenvalues of $\mathbf{C}$ are found by diagonalization of the covariance matrix. The eigenvectors denote the orthogonal modes of motion, and the eigenvectors with the largest eigenvalues dominate the dynamics of the system.

2.4. Supercomputing Grid Infrastructure. In order to perform the study reported here, very substantial computing resources were needed. To this end, three supercomputing grid infrastructures were utilized in the United States and Europe. These include computing nodes at Leeds and Rutherford Appleton Laboratories on the UK's National Grid Service (NGS);[47] Bigben at Pittsburgh Supercomputing Center (PSC), Ranger at the Texas Advanced Computing Center (TACC), and Kraken at the National Institute for Computer Science (NICS), all on the U.S. TeraGrid.[48] RZG at Rechenzentrum Garching on the Distributed European Infrastructure for Supercomputing Applications (DEISA)[49] and Intrepid at the Argonne Leadership Computing Facility were also employed to run these simulations, as well as local machines at UCL, including the Centre for Computational Science's Linux cluster, Mavrino, and UCL Research Computing's Legion machine. Submission of jobs was facilitated by the Application Hosting Environment (AHE),[50] now in its second version.[51,52] AHE allows the submission of geographically distributed applications through a single, uniform interface

(43) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*, 1st ed.; Reidel Publishing Co.: Dordrecht, 1981.
(44) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19.
(45) Thyveetil, M.-A.; Coveney, P. V.; Suter, J. L.; Greenwell, H. C. *Chem. Mater.* **2007**, *19*, 5510–5523.
(46) Jha, S.; Coveney, P. V.; Laughton, C. A. *J. Comput. Chem.* **2005**, *26*, 1617–1627.
(47) National Grid Service, http://www.ngs.ac.uk/.
(48) TeraGrid, http://www.teragrid.org/.
(49) Distributed European Infrastructure for Supercomputing Applications, http://www.deisa.eu/.
(50) Application Hosting Environment, http://www.realitygrid.org/AHE/.
(51) Zasada, S. J.; Coveney, P. V. *Comput. Phys. Commun.* **2009**, *180*, 2513–2525.
(52) Coveney, P. V.; Zasada, S. J.; Saksena, R. S.; McKeown, M.; Pickles, S. *Comput. Phys. Commun.* **2007**, *176*, 406–418.

which interoperates between Globus and Unicore grids, and also retrieves output data automatically to a local machine once the simulation has completed. The total combined CPU time expended in this work was in excess of 4 M CPU hours and resulted in >350 ns of production molecular dynamics. Our simulations produced in excess of 4 terabytes of data which was transferred via the grid infrastructure to local storage at UCL, and backed up using tape storage on the U.S. TeraGrid at the National Centre for Supercomputing Application (NCSA) in Illinois.

Visualization is an essential tool for examining and analyzing the structure and behavior of molecular systems. Our systems were visualized using the AtomEye atomistic configuration viewer[53] and the Visual Molecular Dynamics program (VMD).[54] AtomEye and VMD are both freely available visualization software packages. AtomEye treats atoms as spheres and bonds as cylinders. These images are rendered as primitive objects rather than composites of polygons. AtomEye was employed for its fast rendering qualities. VMD was used for its excellent representation of nucleic acid structures.

3. Results

In this section we discuss the various differences in the interaction of RNA with an aqueous clay surface in the presence of monovalent or divalent cations and discuss how the different cations mediate the interaction between the clay and RNA. We report on the observed folding and formation of secondary structural motifs in the various RNA sequences. We also discuss the mechanism by which RNA tethers to the montmorillonite clay surface and account for the observed increase in adsorption of RNA on the clay surface in the presence of divalent cations.

3.1. Cations as Mediators of the Adsorption of RNA on Clay Surfaces. Model **Ia** was built in order to test the effects of the initial structure and distribution of the counterions on the evolution of the MD simulation. Model **Ia** is identical to model **I** in every aspect except for the initial location of the Na^+ counterions. A comparison of the first 23 ns of simulation from models **I** and **Ia**, as well as average structures, showed that, once equilibrium had been reached, both systems exhibited more or less identical RNA structure and conformation and counterion arrangement; this is supported by the similarities in the density profiles produced for each system (data not shown). Insignificant differences in models **I** and **Ia** can be attributed to the random nature of MD simulations and did not affect the time-averaged data of the two systems.

RNA−montmorillonite systems were simulated with two different choices of cation to balance the negative charge arising due to the unprotonated phosphate groups present in the RNA molecule and from the partial substitutions in the clay sheets. The RNA−clay models were simulated with Na^+ and Ca^{2+} ions for the RNA sequences **A**, **B**, and **C**, as described in Table 2. Previous experimental studies carried out on nucleic acid−mineral systems[10,17] have shown that the cations play a key role in mediating the interaction between RNA and clay surfaces. This role was reported by Franchi et al. using adsorption assay experiments.[10] Franchi and co-workers reported that divalent cations (Ca^{2+} and Mg^{2+}) are more efficient than monovalent cations (Na^+) in mediating the adsorption of nucleic acids on clay minerals including montmorillonite and kaolinite. This effect can be explained by the higher affinity of divalent cations for nucleic acid molecules and their greater ability to counterbalance the negative charges present in the nucleic acid−clay system. Visualization of the average structure taken after 20 ns, shown in Figure 4, displays the extent of the differences caused by the different cations. The RNA molecule interacts very strongly with the surface when the divalent Ca^{2+} ions are present in models **II**, **V**, and **X**, compared to the RNA strand in models **I**, **IV**, and **IX** when monovalent ions are present. Despite the like-charge of the RNA and clay surface, the RNA polyanion is attracted to the surface when bridging divalent cations are available in models **II**, **V**, and **X**. Thus, as in experiments, our simulations show significant differences in the interaction of RNA with the clay surfaces in the presence of Na^+ and Ca^{2+} cations.

Table 2. RNA Sequences and Their Corresponding Optimal Folded Secondary Structures for Sequences **B** and **C** and the Suboptimal Structures in Models **V**, **VI**, **IX**, and **X**; Secondary Structures Are Shown in the "Dot−Bracket" Notation (Explained in the Main Text), along with Their Calculated Free Energies

model	sequence	secondary structure	type	free energy (kcal/mol)
	B	....(((((.....))))).....	optimal	−6.15
V	**B**	..((.............))...	suboptimal	2.51
VI	**B**	.....(.....).....	suboptimal	2.50
	C	...(((((((((...))))))..))).	optimal	−7.40
IX	**C**	.....	unfolded	0.00
X	**C**	...(.....).....	suboptimal	4.20

Plots of atomic density for models **I** and **II**, displayed in Figure 5, show the differences in density of the phosphorus atoms near the clay surface when either Na^+ or Ca^{2+} ions are present. These plots clearly show that more phosphorus atoms in model **II** (Ca^{2+}) are concentrated near the clay surface than in model **I** (Na^+), indicating that RNA has a greater propensity to be closer to the surface in the presence of divalent cations. The mean-squared displacement plots in Figure 6 for models **I**, **II**, **V**, and **VI** show greater diffusive molecular motion for RNA when Na^+ ions are present, as compared to the models containing Ca^{2+} ions. Thus, the mean-squared displacement of RNA as a function of simulation time for models **II** and **V** shows very little diffusive motion, characterized by the gradient of the mean squared displacement plots in Figure 6. This supports the earlier claim that divalent cations are more efficient than monovalent cations in mediating the adsorption of nucleic acids on clay minerals. Franchi's experiments showed no significant differences in the adsorption of polypurines and polypyrimidines on montmorillonite, even though previous work by Ferris and Ertem[55] reported a higher affinity of free purines and purine nucleotides than that of free pyrimidines and pyrimidine nucleotides.

Visualization of the snapshots taken from MD trajectories of the entire system in models **II**, **V**, and **X**, displayed in Figure 7, shows the RNA strand tethering to the surface through a single planar purine base of the biopolymers with sequences **A**, **B**, and **dA**. Model **XIII** shows a Watson−Crick double-helical strand of RNA, sequence **dA** (Table 2), tethered to the montmorillonite surface through a planar purine base at the 5′-end of one of the strands, rather than midsection nucleotide base. Average structures of the RNA molecule, shown in Figure 8, reveal the extent to which the conformation of RNA has altered in the presence of the clay surface, compared to the elongated helical conformation of the strands in aqueous water. Experiments exhibit folding of single-stranded RNAs in bulk water and RNAs on mica surfaces.[18,37]

(53) Li, J. *Modell. Simul. Mater. Sci. Eng.* **2003**, *11*, 173–177.
(54) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
(55) Ferris, J. P.; Ertem, G. *Science* **1992**, *257*, 1387–1389.

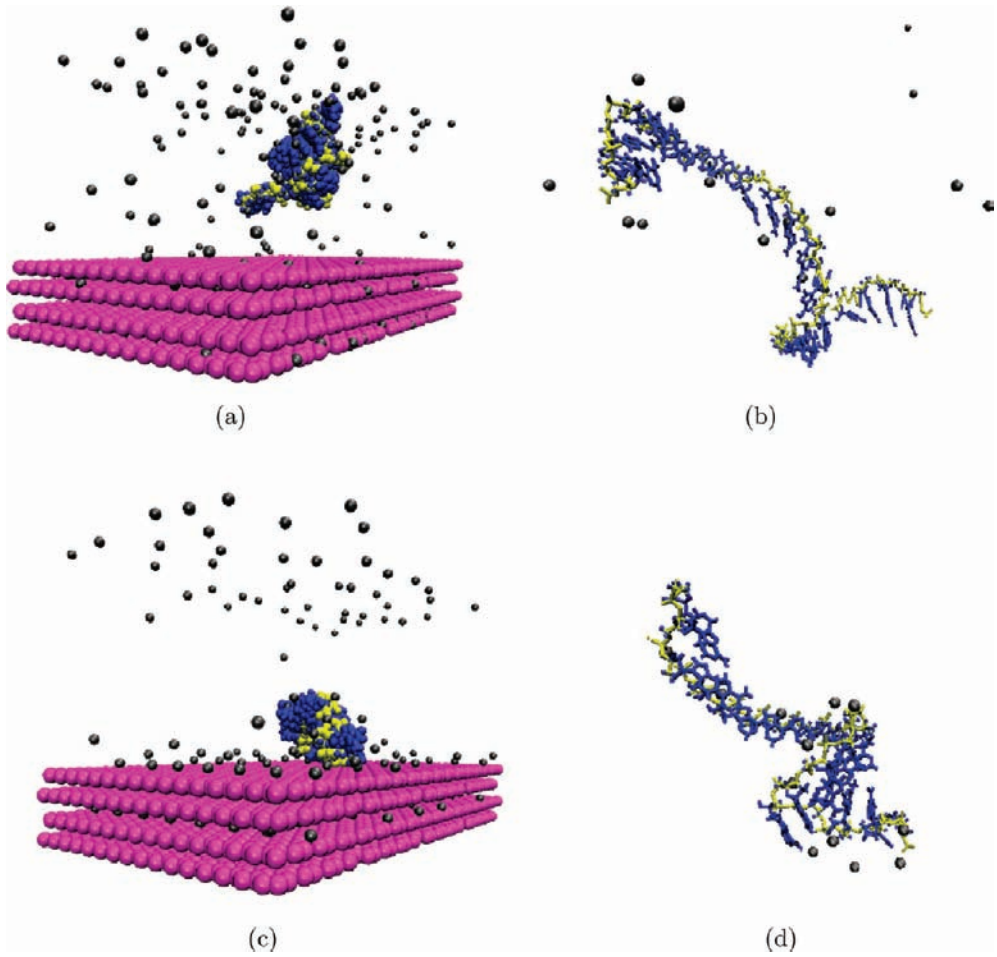

Figure 4. Average structures taken from our simulations after 20 ns of simulation time. Atom colors are as described in Figure 2, water is not shown for clarity. The visualizations of the trajectories here reinforces the differences observed in our simulations between the various models described. Visualizations in (a) and (c) respectively show the RNA strand adopting a more compact conformation in models **I** and **II**, which are composed of an aqueous clay surface, compared to the extended structure of RNA in bulk water, observed in models **III** and **IV** in (b) and (d), respectively. Visualizations displayed here also show the nonadjacent stacking of nucleotide bases in models **I** and **II** (a) and (c), respectively, which is not observed in the bulk water models. Color scheme is the same as in Figure 3.

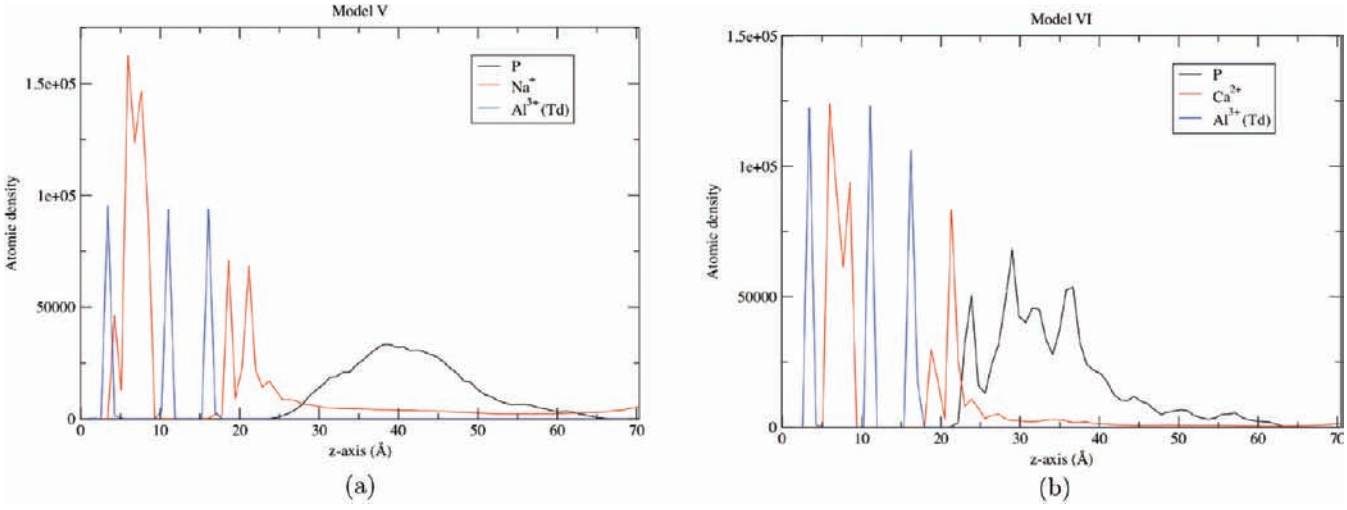

Figure 5. Atomic density profiles taken from 20 ns simulations, of models **V** in (a) and **VI** in (b). The atomic density profile shows the time averaged atomic density of the specified atoms as a function of distance along the *z*-axis of the simulation cell. Atomic density refers to the number of atoms per unit volume. These profiles show the density of phosphorus atoms belonging to the RNA molecules, surface Al^{3+} ions and Na^{+} or Ca^{2+} cations. The plots show the density of phosphorus atoms to be concentrated closer to the surface in model **VI** than in model **V** where the density is spread-out around 40 Å. The density of phosphorus atoms in model **VI** shows more well-defined peaks, indicating that these atoms occupy more regions relative to the clay surface than those in model **V**. These differences can be attributed to the increased electrostatic screening by divalent Ca^{2+} ions in model **VI** relative to that of monovalent Na^{+} ions in model **V** in (a).

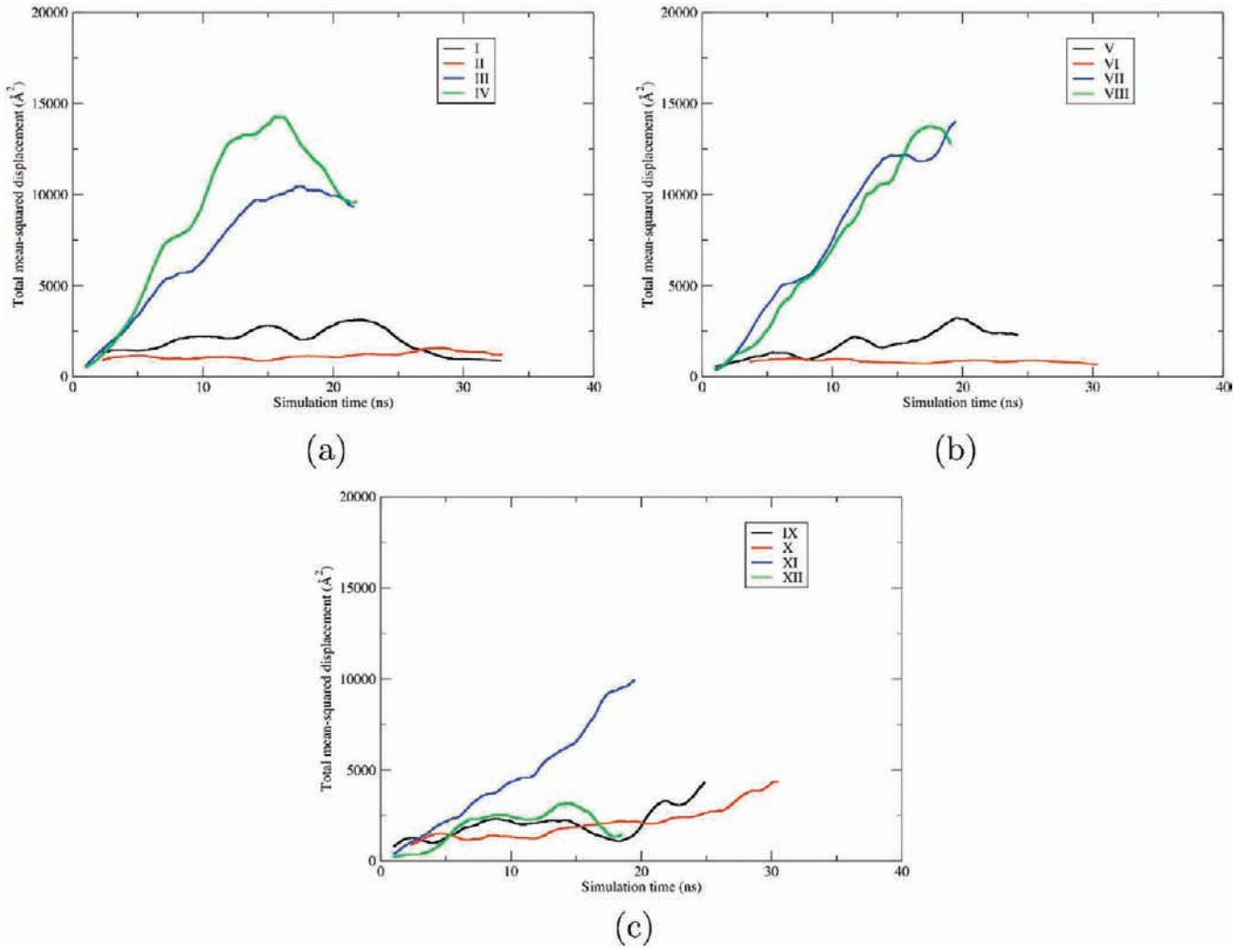

Figure 6. (a) Total mean-squared displacement of RNA for sequence **A**, systems **I**−**IV**, (b) sequence **B**, **V**−**VIII**, and (c) sequence **C**, **IX**−**XII**, all at ambient temperature and pressure (300 K, 1 atm), relative to the initial structure, as a function of simulation time. The linear portion of the MSD plot is proportional to the diffusion coefficient of the diffusing atoms. The plots show a significant difference in the diffusion of RNA in an aqueous solution (green and blue) as compared to RNA interacting with a clay surface (black and red). When interacting with the clay surface and calcium ions (systems **II**, **VI**, and **X**) the RNA strand exhibits little diffusive molecular motion, characterized by the low gradient of the slope. This is due to the increased screening of negative charges from phosphate groups and surface sites by divalent calcium ions, allowing RNA to become attracted to the surface cooperatively, bringing the strand in close proximity to the surface and permitting it to tether through a nucleotide base. The RNA strand combined with the clay surface and sodium ions (system **I**, **V**, and **IX**) displays more diffuse motion than with Ca^{2+} ions due to the reduced screening of negative charges by Na^{+} ions. This is observed in the gradient of the MSD plots (black line). The MSD plots for RNA in bulk water, models **III**, **IV**, **VII**, and **VIII**, show interesting step-function-like behavior. The RNA-in-bulk-water models (blue and green curves) show a decrease in diffusivity between ∼5−10 ns for a duration of 2.5 ns, which then return to their original behavior.

Our simulations show that the nucleotide base contributes to the mechanism by which RNA adsorbs onto the aqueous montmorillonite surface. This disagrees with the notion that the sugar and base groups of the RNA molecule do not take part in the adsorption of RNA onto the surface and that the negatively charged phosphate groups are the best candidates for the interaction with clay surfaces mediated through cations, suggested by Franchi et al.[10] The visualizations from our simulations in Figure 7 show the planar adenosine base lying flat on the montmorillonite surface at a tetrahedral Si^{4+} site, away from the Al^{3+} partial substitutions. The adenine ring interacts with the clay surface atoms via a combination of van der Waals and electrostatic forces. The experiment of Franchi and co-workers indicates that double-stranded nucleic acids need higher cation concentrations than single-stranded ones to establish an interaction with the clay surface. One proposed mechanism by which nucleic acids interact with a montmorillonite surface is through a "cation bridge" where divalent cations locate between the mineral surface and the nucleic acid; the nucleic acid is then cooperatively bound to the surface through the divalent cations.[10] Our simulations do not exhibit the binding of RNA through a "cation bridge" mechanism, but rather indicate that the divalent cations neutralize the negative charge present on both the nucleic acid and the clay surface. The RNA then binds to the surface though a single nucleotide base essentially mediated by Lennard-Jones dispersion forces.

Pastré et al. used atomic force microscopy (AFM) to probe DNA adsorbed to a mica surface.[17] Their experiments suggest that DNA attraction is due to the sharing of the DNA and mica counterions. The correlation between divalent counterions on both the negatively charged DNA and the mica surface can generate a net attractive force, whereas the correlation between monovalent counterions is relatively

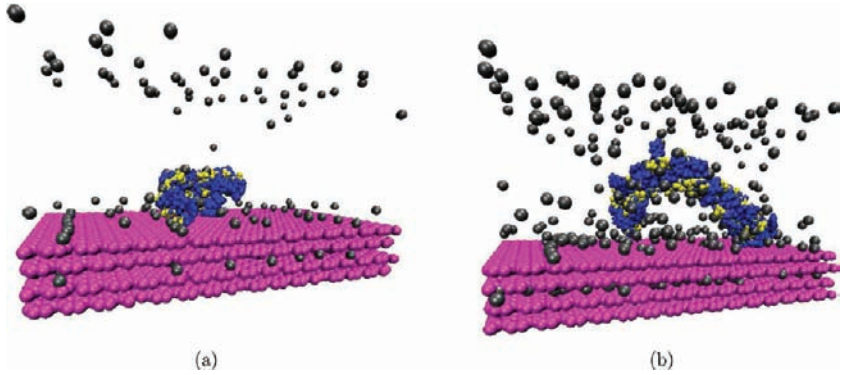

Figure 7. Simulation snapshots taken from simulations of (a) model **VI** and (b) model **XIII** after 30 and 20 ns, respectively. The color scheme is the same as in Figure 4. These visualizations show the conformation of RNA and throw light on the mechanism by which RNA tethers to the surface. The visualizations in this figure show the tethering of (a) the single strand, of sequence **B**, and (b) the double strand **dA** of RNA to the montmorillonite clay surface, through a planar purine ring, in this case adenosine, at the 5′-end of the biopolymer. The screening of the negative charges between the surface and phosphate groups by the cations allows the RNA strand to get close enough to the surface for it to tether through the nucleotide base and the surface via attractive dispersion forces. Water molecules are hidden to aid in viewing.

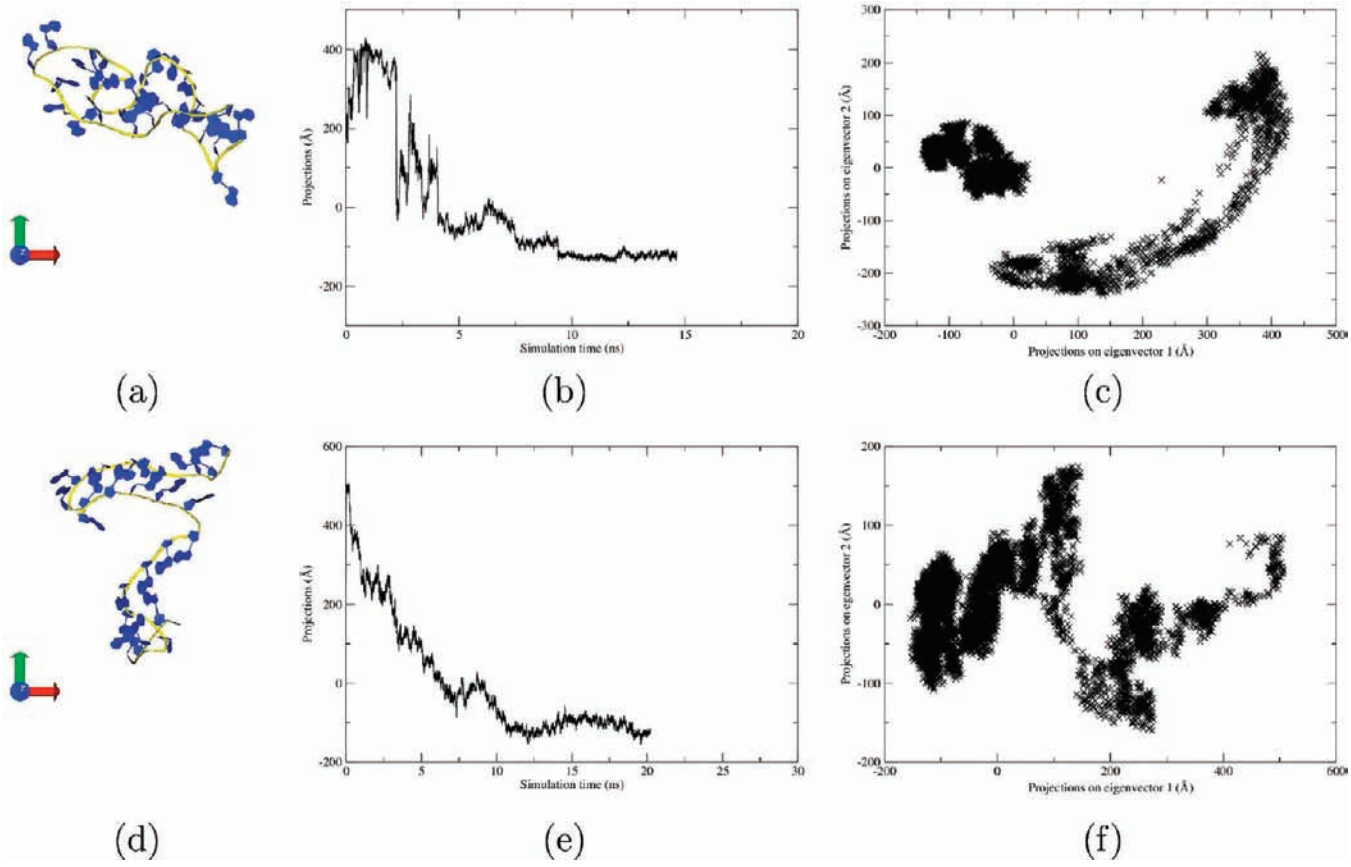

Figure 8. PCA of RNA interacting with a clay surface used to elucidate the principal modes of motion within systems **I**−**IV** of sequence **A**. (a,d) Average structure of RNA in models **I** and **II**, respectively, averaged over 20 ns of simulation. (b,e) Plots of projections against simulation time, while (c,f) are plots of projections of mode 1 against projections of mode 2. The motions along the first eigenvector obtained from coordinates of the RNA molecule and their planes of motion are defined by the first two eigenvectors of model **I** in (a−c) and model **II** in (d−f). The color scheme is the same as that in Figure 1. Both models show a strong dependence on the first two eigenvectors, indicating that forces acting along these eigenvectors are coupled. Adding the clay surface and tethering of the RNA significantly restrict the movement of the RNA molecule. The main contribution to the RNA dynamics comes from the ends of the molecule.

ineffective. These experimental observations regarding the counterion valency correlate with the simulations reported here in which Na^+ is ineffective in binding the *ss*RNA in models **I**, **VI**, and **X**. The degree to which the Na^+ ions are ineffective at attracting the RNA to the montmorillonite surface can be viewed qualitatively in the animation of model **VI** supplied in the Supporting Information.

3.2. RNA Folding and Formation of Secondary Structural Motifs. RNA molecules are known to fold into specific, intricate, three-dimensional structures which are necessary to perform

numerous biological functions.[56] Thus, understanding RNA folding is important in unravelling RNA's putative role in the early chemical processes which may have led to the origin of life on Earth. In addition to the models consisting of RNA on an aqueous montmorillonite surface, corresponding models of RNA in bulk water were built and simulated, as discussed in section 2.1. These are models **III**, **IV** for sequence **A**, **VII**, **VIII** for sequence **B**, and models **XI**, **XII** for sequence **C**; see Table 2. The bulk water simulations were designed to understand how the structure and conformation of different RNA sequences are modified in the presence of a mineral surface, by comparison with their corresponding behavior in bulk aqueous solution. Visualizations taken from our simulations reveal significant changes in the conformation of RNA between the clay surface models **I** and **II** and the bulk water models **III** and **IV**, as well as notable differences in structure and conformation when different cations are present in models **I** and **II**.

The first experiments on the initial collapse of the RNA *Tetrahymena* ribozyme, induced by various counterions, were presented by Thirumalai and Heilman in 2001.[57,58] Experiments using time-resolved small-angle X-ray scattering (trSAXS) have revealed that, upon addition of divalent Mg^{2+} cations, RNA in bulk water rapidly compacts from an extended state to a globular state of nearly native dimensions, well before the formation of any stable tertiary contacts.[56] Gel electrophoresis experiments on the *Tetrahymena* ribozyme performed by Kucoli et al. show that the rate of formation of the catalytically active native state, from the initial collapsed state, increases with decreasing counterion charge.[59] It is intuitive that the transition to the active state would be slower compared to the initial collapse as it requires breaking of hydrogen bonds and refolding. These collapsed RNA states can be considered analogous to the molten globule states that appear in early stages of protein folding which are stabilized by hydrophobic interactions. Our simulations show a collapse of the initial elongated RNA structures, for sequences **A**, **B**, and **C** shown in Figure 1, into folded, more compact structures in the models in which RNA is interacting with a montmorillonite surface (models **I**, **II**, **V**, and **VI**), over a time scale of 5 ns. The collapse of the initial structure can be seen in plots of radius of gyration of the RNA strand as a function of simulation time, shown in figure 9. The radius of gyration is calculated here for the RNA molecule in each model. The radius of gyration plots in Figure 9 for models **I** and **II** show a decrease in the effective size of the molecule over the first 15 ns of simulation, from 20 to around 15 Å. The radius of gyration plots for models **V** and **VI** show a decrease in the effective size from 25 to around 15 Å. The reduction in size of the RNA single-strand depicted here is supported further by the observed folding of RNA seen during our simulations. The 5 ns collapse observed in our simulations is analogous to the 1 μs fast collapse of the *Tetrahymena* ribozyme reported experimentally using trSAXS.[56] The initial folding of the ribozyme happens over a longer time scale due to the size (400-mer) and complexity of the ribozyme compared with those of the smaller RNA sequences studied here. The ribozyme has complex junctions stabilized by hydrogen bonds and base-stacking interactions that are compatible with the RNA's electrostatically extended conformation. Base stacking refers to the stabilizing effect of parallel nucleotide base groups that interact through the π-bonding orbitals of their aromatic rings. The breaking of these initial interactions slows down the rate of the earliest global conformational changes. In the *Tetrahymena* ribozyme experiments discussed here, the ribozyme's initial "extended" structure already contains preformed secondary structures; it is not a random elongated strand.

The folding that is observed in models **I**, **II**, **V**, **VI**, **IX**, and **X** containing clay surfaces is not observed in the corresponding bulk water simulations in this study. The single strand RNA sequences in models **III**, **IV**, **VII**, **VIII**, **XI**, and **XII** retain their initial helical structure over the course of each simulation. The absence of folding of RNA in these bulk water simulations is characterized by the constant effective size of the RNA molecule for the duration of each simulation, which can be seen in plots of the radius of gyration in Figure 9.

The outcome of the initial collapse of RNA has been suggested to result from a trade-off between the electrostatic screening of negatively charged unprotonated phosphate groups versus the hydrophobic collapse of the strand to shield the hydrophobic base groups.[56] The folding of the initial structures in our simulations leads to the formation of secondary structural motifs in models **I**, **II**, **V**, and **VI**. The formation of the stem−loop secondary structure, shown in Figure 8, is stabilized by the stacking of adjacent and nonadjacent bases, and by the formation of Watson−Crick base pairs in models **V** and **VI**. The occurrence of Watson−Crick base-pairs and base stacking are known to stabilize the formation of optimal and suboptimal folded secondary structures.[26]

To analyze the principal modes of motion of RNA in all of our models, PCA was employed, as described in section 2.3; our results are shown in Figure 8. In published MD simulations of RNA in bulk water the dominant modes of motion are a twisting of the strand around a central axis, junction bending, and wedge bending.[26] Models **III**, **IV**, **VII**, **VIII**, **XI**, and **XII**, which consist of RNA in bulk water, confirm these findings. The dominant modes of motion of RNA in bulk water can be seen in Figure 8. The clay-surface interacting RNA in models **I**, **II**, **V**, **VI**, **IX**, and **X** have different modes of motion since it is the interaction with the clay surface which restricts molecular movement. The clay/RNA models show a strong dependence between the first two eigenvectors, suggesting that forces acting along these eigenvectors are coupled. The presence of the clay surface and the resultant tethering of the RNA significantly restricts the movement of the RNA molecule. The main contribution to the dynamics is seen at the ends of each of the RNA oligomer chains, analogous to our previous simulations of DNA intercalated within a layered double hydroxide (LDH) matrix.[21]

The compact folded structures observed in the simulations are analogous to the collapsed, folded, and misfolded hairpin loops observed during replica-exchange molecular dynamics performed by Sorin et al.,[60] who reported the stabilization of misfolds by non-native interstrand nucleotide stacking. Sorin et al. employed large ensembles of simulations using the *Folding@home*-distributed computing grid, thereby showing the competing pathways in hairpin folding and unfolding. Sorin

(56) Das, R.; Kwok, L. W.; Millett, I. S.; Bai, Y.; Mills, T. T.; Jacob, J.; Maskel, G. S.; Seifert, S.; Mochrie, S. G. J.; Thiyagarajan, P.; Doniach, S.; Pollack, L.; Herschlag, D. *J. Mol. Biol.* **2003**, *332*, 311–319.
(57) Thirumalai, D.; Lee, N.; Woodson, S.; Klimov, D. *Annu. Rev. Phys. Chem.* **2001**, *52*, 751–762.
(58) Heilman-Miller, S.; Pan, J.; Thirumalai, D.; Woodson, S. *J. Mol. Biol.* **2001**, *309*, 57–68.
(59) Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. *J. Am. Chem. Soc.* **2007**, *129*, 2676–2682.
(60) Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. *Biophys. J.* **2003**, *85*, 790–803.

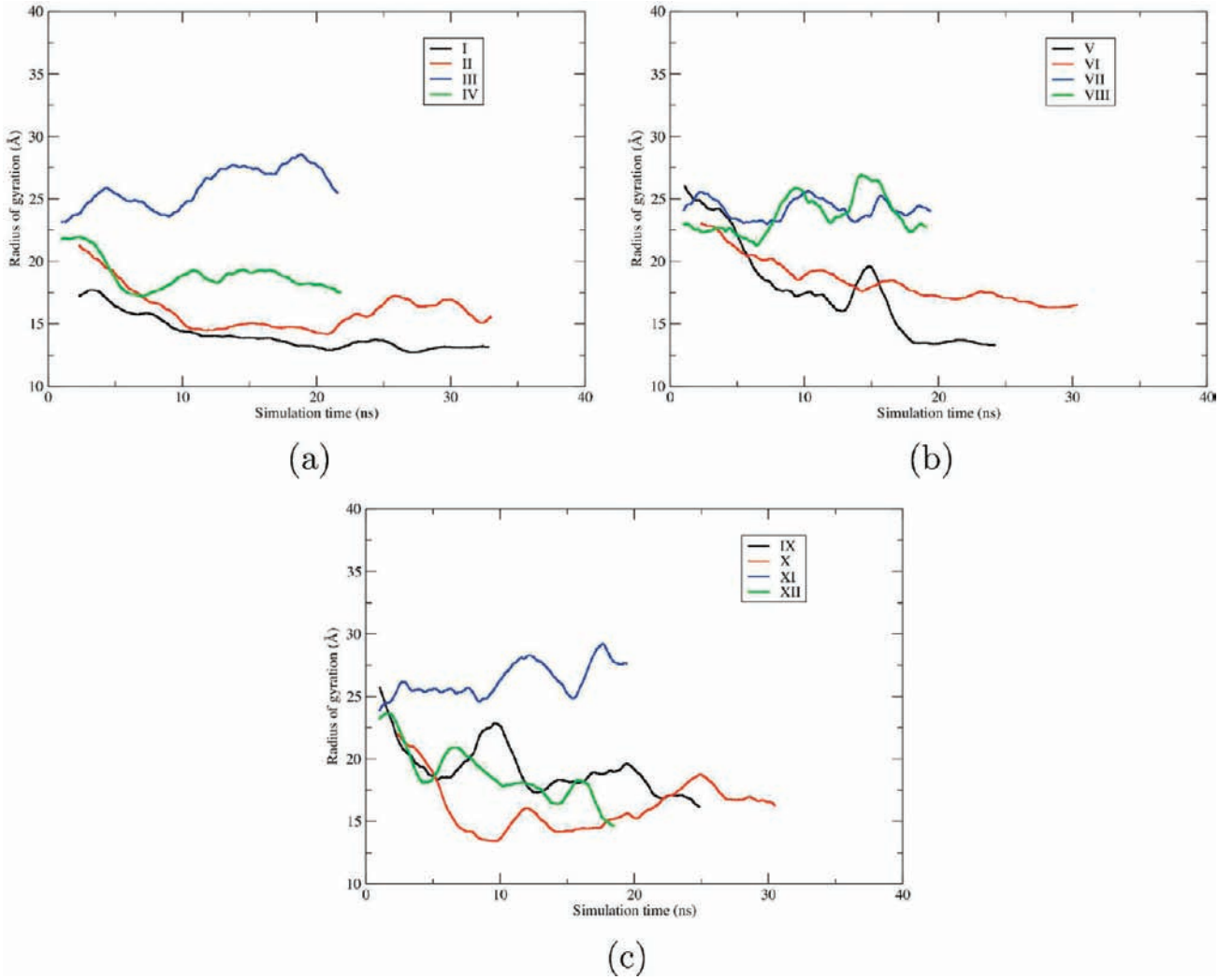

Figure 9. Radius of gyration, r_G, of RNA as a function of simulation time, relative to the initial structure for (a) models **I**−**IV**, (b) models **V**−**VIII**, and (c) **IX**−**XII**. Lines in black and red indicate models with a clay surface, while blue and green lines refer to models in bulk water. The radius of gyration is a measure of the effective size of the RNA sequence. The radius of gyration is significantly smaller in the case of RNAs interacting with the montmorillonite surface, most likely due to the increased concentration of cations near the surface of the clay more effectively neutralizing the charge of the phosphate groups, reducing the electrostatic penalty of having a compact structure and encouraging a folded conformation, stabilized by the "burying" of hydrophobic nucleotide bases and the formation of adjacent and nonadjacent base stacking, as well as Watson−Crick base pairing. The concentration of ions in the models of RNA in bulk aqueous solution (blue and green) is less than in the clay surface systems; as a result the RNA favors a conformation with a greater effective size, which corresponds to an elongated helical structure close to that in the initial models.

et al. concluded that the conformational dynamics for folding small 12-mer RNA sequences happens over tens of milliseconds. Our models indicate that the 25-mer RNA, of sequences **A**, **B**, and **C**, in the presence of a clay surface, enjoy an increased rate of folding due to the change in their modes of motion.

For sequences **B** and **C**, the optimal and lowest-energy folded structures were calculated using the *RNAfold* tool which is part of the Vienna RNA software suite.[31] The *RNAfold* tool reads in RNA sequences and calculates their minimum free energies using the minimum free energy algorithm of Zuker and Stiegler.[61] The optimal folded structures of sequences **B** and **C** are displayed schematically, and as a three-dimensional structure, in Figure 3. The optimal folded structures and energies were used to compare against the folded structures of the RNA sequences in our simulations, shown in Figure 11. To obtain free energies for the folded structures that arise in our simulations, the *RNAinverse* program, which is also part of the Vienna suite, was used. The *RNAinverse* program searches for sequences that map into a predefined tertiary structure, thereby inverting the folding algorithm. Target structures and starting sequences are read-in alternately. The optimal and suboptimal structures and energies are tabulated in Table 2 where the secondary structures have been displayed using dot−bracket notation in which the characters "(" and ")" correspond to the 5′ and the 3′ base which close the stem−loop motif, while "." denotes an unpaired base. The RNA-folding pathway can be discerned from the free energy landscape which may be visualized in terms of a tree diagram. The tree arranges local minima and their corresponding saddle points in a hierarchical fashion. Figure 12 shows the tree diagrams for both sequences **B** and **C**. The optimal fold, the lowest-energy fold, which has the structure shown in Figure 2, is given the label 1 in the tree plots.

The folding landscape in Figure 12 shows the folded structures for sequences **B** and **C** within the folding pathway to their optimal lowest-energy fold in bulk aqueous solution. The structures taken from our simulations are suboptimal

(61) Zuker, M.; Stiegler, P. *Nucleic Acids Res.* **1981**, *9*, 133–148.

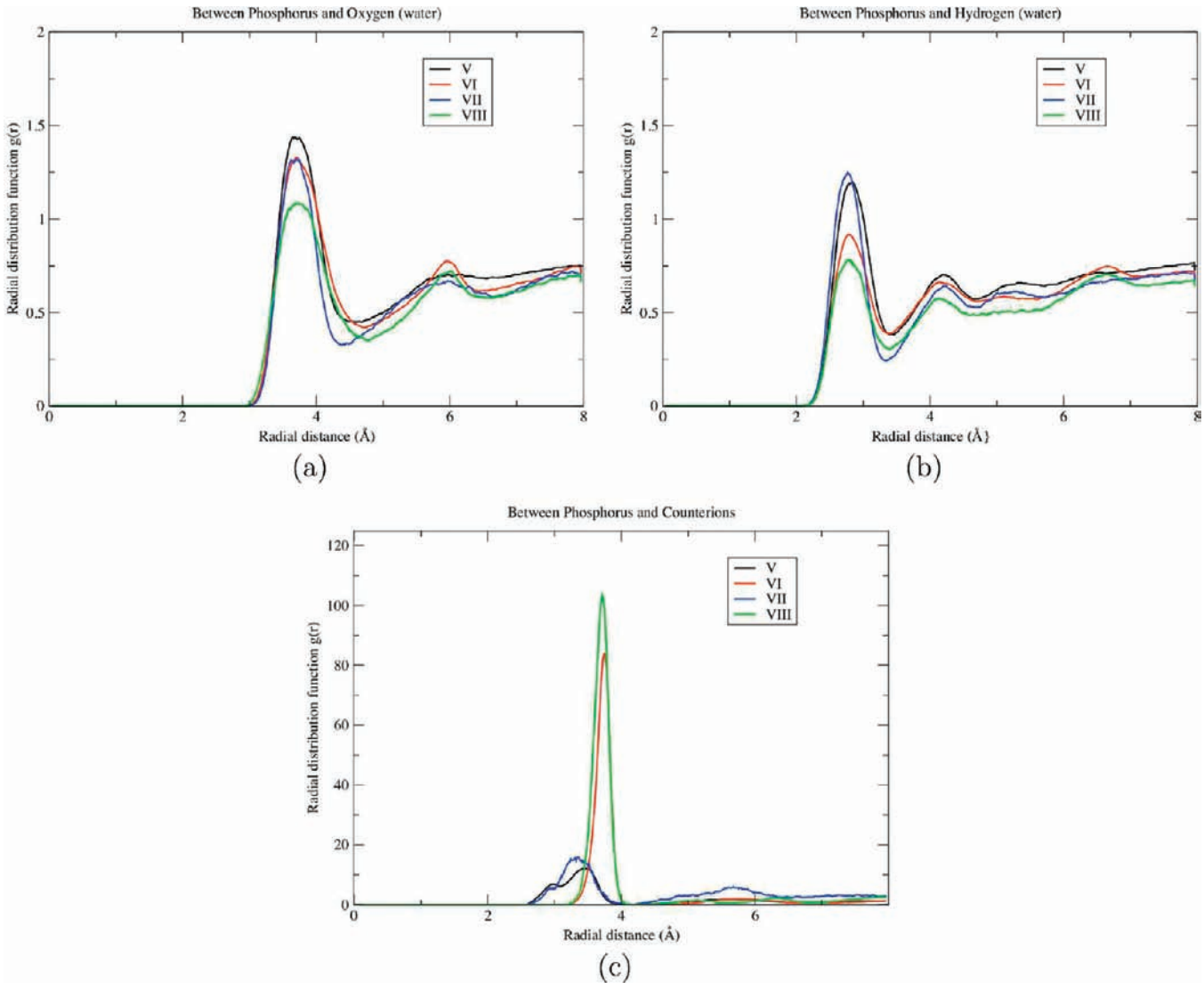

Figure 10. Average radial distribution functions for all the models which include sequence **C** with a clay surface and in bulk water (models **V**, **VI**, **VII**, and **VIII**), calculated with respect to phosphorus atoms in phosphate groups of RNA molecules. Radial distribution functions show similar trends for coordination of water molecules around a central phosphorus atom. The radial distribution functions show (a) water oxygen atoms at 4 Å, (b) water hydrogen atoms localized at 2.8 Å and (c) counterions with respect to phosphorus, where the cations are Na^+ (models **V** and **VII**) and Ca^{2+} (models **VI** and **VIII**). Significant difference in the radial distribution for the two different cations. That for Na^+ ions exhibits a "shoulder" on the main peak at 3.4 Å, indicating that Na^+ ions occupy two sites around phosphorus atoms at 3 Å and 3.4 Å. The Ca^{2+} case by contrast only has one site at which the ions are located.

structures which are all present along the predicted folding pathway. This indicates that these local minima are occupied at intermediate stages in the folding of the optimal secondary structure conformations. As folding is not observed after 20 ns of simulation in our bulk water models **VII**, **VIII**, **XI**, and **XII**, but folded conformations of suboptimal energies are observed along folding pathways for our systems containing RNA and a clay surface in models **V**, **VI**, **IX**, and **X**, we infer that the clay surface has the effect of increasing the rate at which these RNA sequences compact and fold along pathways that might eventually reach their optimally folded conformations, associated with global free energy minima, even though these may not be found in a clay/RNA system due to the presence of many additional interactions.

Ion size determines the closest distance between ion and RNA, affects ion−ion distance and hence the strength of Coulombic and excluded volume correlation. Folding studies have been performed on the *Tetrahymena* ribozyme looking into the effects of differing ions on folding kinetics.[62] These experiments show that there is a linear relationship between increase in folding stability and the increase in (divalent) ion charge density (i.e., charge/volume of the ion). This experimental observation supports the differences in folding rates shown between the simulation of models **I**, **V**, and **IX** with models **II**, **VI**, and **X**. The increase in charge density of Ca^{2+} over Na^+ leads to increased folding stability and drives the increase in rate observed in these simulations.

During the full 25 ns MD simulation of model **XIII** the double strand of RNA does not undergo any folding, but retains its double helical conformation; this indicates that the Watson−Crick base pairing is stable under these conditions, and no unzipping of one strand from its complementary strand is observed. Figure 7 shows the *ds*RNA tethering to the montmorillonite surface through an adenine base group at the

(62) Koculi, E.; Lee, N. K.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2004**, *341*, 27–36.

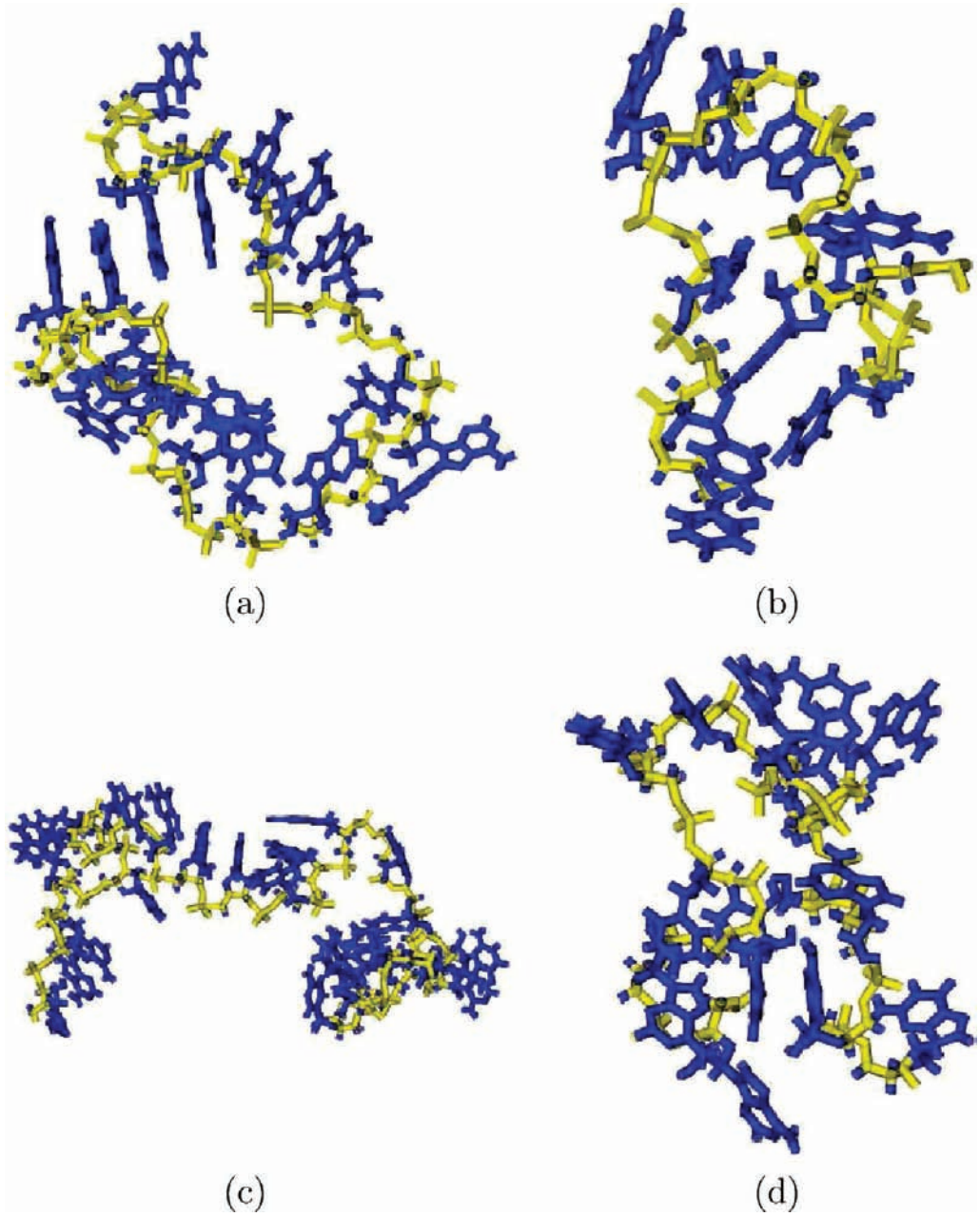

Figure 11. Folded sections of the average RNA structure taken from the entire production simulation for models (a) **V**, (b) **VI**, (c) **IX**, and (d) **X**. Color coding as in Figure 1. Images show the folded secondary structures of RNA for sequences **B** and **C** when interacting with a clay surface. (a, b, d) Formations of hairpin loop-type motifs stabilized by stacking of nucleotide bases, seen as a closed loop in the 3D structures. (c) Sequence **C** in model **IX** which has not yet formed a secondary structure motif but exhibits a more compact conformation than the corresponding sequence in bulk water in model **XI**, in which the end segments have folded in onto the rest of the molecule.

5′-end of the tethered strand; this suggests that a tethered strand provides a stable template for growth of a complementary strand in the 5′ to 3′ direction.

4. Conclusion

In this present study we have elucidated a number of facets of nucleic acid-clay interaction and reactivity that have been observed in previous experimental studies, but have hitherto remained unexplained at a molecular level.

One prevailing theory in origins of life studies is that RNA in prebiotic conditions may have been confined and restricted in motion by mineral surfaces, much as it is confined within the ribosome of living cells today.[14] Due to the ubiquitous presence of clay minerals and their associated cations in nature, these aqueous environments may have provided prebiotic habitats for nascent nucleic acid polymers. Indeed, clays may thus have played a key role in the formation and preservation of these polymers and/or their precursors. Analysis of the dynamics of the RNA oligomers in all of the models we have studied reveals that the behavior of RNA interacting with these surfaces is significantly different from that in bulk water.

In general, we have shown that strong electrostatic forces act between the montmorillonite clay surface, the interacting RNA molecule and the aqueous monovalent or divalent cations. Divalent cations are more effective in mediating the interaction between the clay surface and RNA sequences,

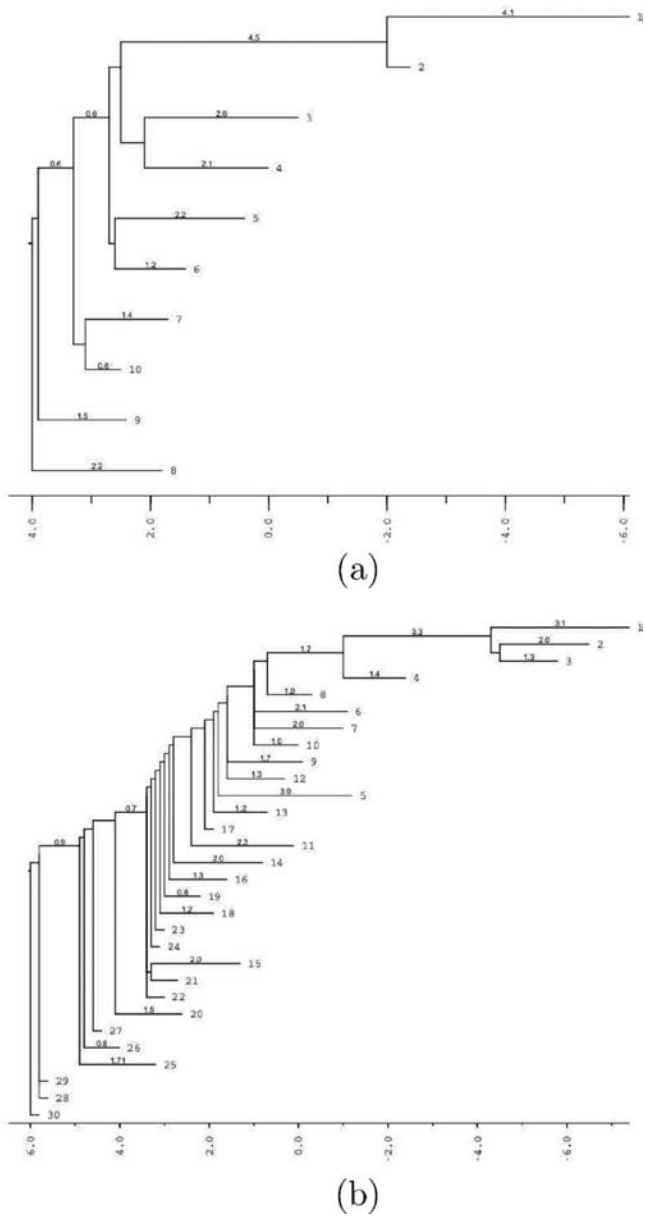

Figure 12. Representation of the folding free energy landscape of (a) sequence **B** and (b) sequence **C**. The tree arranges local minima and their corresponding saddle points in a hierarchical fashion. The optimal and lowest free energy fold is labeled as 1 in each tree. Folding landscapes were calculated using the Vienna RNA software suite.[31] Energies calculated for each RNA structure in figure 11 correspond to suboptimal folds and energies in the folding landscapes shown here. The energy values given in the horizontal axis are in units of kcal/mol. The vertical axis shows the relative stability of the folded structure, the highest being the most stable. The free energies of the RNA sequences in models **V** and **VI** correspond to the suboptimal structures labeled 9 and 10, respectively, in (a). Model **IX** corresponds to the suboptimal structure labeled 27 in (b).

due to increased electrostatic screening. The electrostatic repulsion of the unprotonated phosphate groups within the RNA molecule from the negatively charged sites associated with aluminum ions on the clay surface is overcome through the presence of solvated divalent cations, allowing the RNA molecule close proximity with the clay surface. We have identifed a mechanism for clay−RNA interactions, whereby a planar purine nucleotide base interacts with the clay surface mainly through Lennard-Jones forces. The electrostatic attraction and tethering of the RNA strand to the clay surface causes it to be significantly "pinned" and hence restricted in movement compared to RNA in bulk water. Structural analysis of surface-interacting RNA molecules in all of the models we have studied confirms that the motion of RNA is indeed substantially restricted compared to RNA in bulk water. Tethering of *ss*RNA to the montmorillonite surface is only observed in models with Ca^{2+} counterions: equivalent models which have Na^{+} counterions do not produce similar effects. This leads to the inference that divalent ions such as Ca^{2+} very likely play a key role in tethering, as reported experimentally.[10] The observed tethering of *ss*RNA to the clay surface is of considerable relevance to prebiotic chemistry: 5′−3′ regioselective RNA polymerization has been reported by Ferris and co-workers,[6−8] but was hitherto difficult to explain mechanistically.

Our simulations show that single strands of RNA fold at a considerably enhanced rate when a montmorillonite surface is present compared with the same sequences in bulk water. For both Na^{+} and Ca^{2+} containing clay and *ss*RNA models, the *ss*RNA undergoes a fast collapse stage of folding into a more compact structure in less than 5 ns, which is not observed in bulk water with the same cations. The increased rate of folding at clay surfaces is of importance to the RNA world view, according to which early RNA polymers would have needed to produce well-defined folded structures to support enzymatic activity in order to cleave self-replicating RNA strands, essential components in the first steps of molecular evolution.[63] Accelerated folding kinetics considerably increases the parameter space that may be explored by clay−RNA complexes in constructing catalytically active structures under conditions which may have been far from benign.

When interacting with the clay surface the motion of the end groups of the oligonucleotide dominate the motion of the molecule; as a result the single RNA strands collapse and fold into more compact structures, made possible by the increased concentration of cations screening the mutual electrostatic repulsion from phosphate groups in clay-containing systems with both Na^{+} or Ca^{2+}. Our simulations show that the compact, folded structures of these RNA sequences are stabilized by the stacking of adjacent and nonadjacent base groups as well as via hydrogen bonds between nucleotide groups, giving rise to secondary structural hairpin loop motifs. Conceptually, once RNA chains are stabilized on clay surfaces, *ds*RNA may be formed, for example as part of a template-directed replication process; folded into tertiary structures, such ribozymal RNA can then act as a catalyst[63] for self-replication.

The insight gained from large-scale molecular simulations performed for tens of nanoseconds on mineral-bound nucleic acids adds considerably to the interpretation of experimental studies through provision of highly detailed molecular level information. It is evident that minerals may substantially modify the chemistries that can occur in RNA world scenarios: an RNA−mineral−water world may have differed considerably from a mineral−free RNA world.

Acknowledgment. This research was supported in part by the National Science Foundation through TeraGrid resources provided by Pittsburgh Supercomputing Center, Texas Advanced Computing Center and National Institute of Computer Science within allocations under NRAC Grant MCA04N014 and MRAC

(63) Wattis, J. A. D.; Coveney, P. V. *J. Phys. Chem. C* **1999**, *103*, 4231–4250.

Grant DMR070013N. We are also indebted to the UK National Grid Service for resources at the University of Leeds and at the Rutherford Appleton laboratory, the DEISA consortium (co-funded by EU FP7 project 508830) for allocations within the DEISA Extreme Computing Initiative at Rechenzentrum Garching, and UCL Research Computing for use of the Legion computer. We thank Tom Cheatham III for in depth discussion on nucleic acid simulation and the Amber forcefield. J.B.S. is funded by an EPSRC PhD studentship award.

Supporting Information Available: MPEG animations of the full MD trajectories for models **V**−**VIII**; the breakdown of the potential energy into its components for system **I**; the change in simulation cell volume as a function of time for model **I**; and the root mean squared deviation of RNA from its initial structure for model **I**. This material is available free of charge via the Internet at http://pubs.acs.org.

JA104106Y