

# Stability of free and mineral-protected nucleic acids: Implications for the RNA world

Jacob B. Swadling, Peter V. Coveney\*, H. Christopher Greenwell<sup>1</sup>

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

Received 15 September 2011; accepted in revised form 13 December 2011; available online 3 January 2012

## Abstract

Using molecular dynamics simulations we study the structural stability of three different nucleic acids intercalated within a magnesium aluminium layered double hydroxide (LDH) mineral, at varying degrees of hydration, and free in aqueous solution. The nucleotides investigated are ribose nucleic acid (RNA), deoxyribose nucleic acid (DNA) and peptide nucleic acid (PNA), all in duplex form. Our simulations show that DNA has enhanced Watson–Crick hydrogen-bonding when intercalated within the LDH clay interlayers, compared with intercalated RNA and PNA, whilst the reverse trend is found for the nucleic acids in bulk water. The tendency for LDH to alter the stability of the three nucleic acids persists for higher temperature and pressure conditions. The uncharged protein backbone of PNA is found to have a detrimental effect on the overall stability of the duplex, as it experiences a greatly reduced electrostatic interaction with the charged LDH sheets compared to RNA and DNA. Assuming an RNA world, in which RNA preceded the DNA/protein world, at some point in time DNA must have taken over the role as the information storage molecule from RNA. These results suggest that a mineral based origin of life may have favoured DNA as the information-storage biomolecule over potentially competing RNA and PNA, providing a route to modern biology from the RNA world.

© 2011 Elsevier Ltd. All rights reserved.

## 1. INTRODUCTION

Identifying the nature of the chemical processes that led to the origin of life on Earth poses a major intellectual challenge to modern science. The complete absence of a “fossil record” for the type of proto-biomolecules present in this era means that there is no way any hypothesis can be proved to be correct. However, it is possible to make reasonable proposals and then experimentally investigate the processes that may have led to the emergence of the first biomolecules, including the first genetic material and the first living organisms (Follmann and Brownson, 2009; Budin and Szostak, 2010).

The RNA world view is one hypothesis that has attracted widespread support amongst those investigating the origin of life (Woese, 1967). The RNA world hypothesis asserts that the first life forms on Earth were based on RNA molecules, which were able to self-replicate, preceding the appearance of proteins and subsequent protein-based (bio)chemistry (Woese, 1967; Joyce, 2002; Orgel, 2004). The essence of the RNA world hypothesis is that the transfer of genetic information relied solely upon the self-replication of RNA, for which Watson–Crick base-pairing provided the key mechanism (Cheng and Unrau, 2010). The hypothesis concurs with the doctrine of biopoiesis, which asserts that life originated from simple organic molecules already present on Earth (Wu and Higgs, 2009). However, a gap exists in the process of biopoiesis in that it relies on there being a pre-biological route to an initial source of nucleotides for the formation of RNA. Though there is evidence showing that RNA could have catalyzed most of the steps involved in the synthesis of nucleotides

\* Corresponding author. Fax: +44 (0)20 7679 4603.

*E-mail address:* [p.v.coveney@ucl.ac.uk](mailto:p.v.coveney@ucl.ac.uk) (P.V. Coveney).

<sup>1</sup> Present address: Department of Earth Sciences, Durham University, South Road, Durham DH1 3LE, United Kingdom.

(Unrau and Bartel, 1998), and possibly the coupling of redox reactions to the synthesis of phosphodiester and peptides, no complete abiotic synthetic pathways for all of the nucleotides have hitherto been reported. Recent work has suggested potentially plausible prebiotic routes to RNA synthesis, including from activated pyrimidine ribonucleotides (Powner et al., 2009; Sutherland, 2010), whilst layered minerals have also been shown to be catalytically active in RNA elongation (Joshi et al., 2009). Notwithstanding these problems, it is conjectured that the transition to the DNA genetic apparatus came at a later time, possibly via an RNA-protein intermediate stage (Jeffares et al., 1998; Poole et al., 1998).

Although it is conceivable that the first forms of living organism were comprised of RNA, oligomers of RNA are highly sensitive to hydrolysis, making it challenging to explain how such species may have survived long enough to create more complex forms of proto-life. Moreover, it has been proposed that RNA's arrival on the scene may have been preceded by simpler organic genetic materials. One such molecule is peptide nucleic acid (PNA) which differs from RNA and DNA in its backbone structure; PNA has an uncharged protein backbone unlike the charged phosphorus backbone of RNA and DNA (Scott et al., 1996) (see Fig. 1). Unlike RNA and DNA, PNA has not been observed in nature; PNA was originally designed using computer modelling and subsequently synthesized in a laboratory.

Experiments have shown that PNA is able to displace complementary oligonucleotide strands from existing DNA duplexes, indicating that PNA-DNA hybrids are more stable than duplex DNA. Other enhanced structural features include faster rates of reformation of the duplex upon denaturation and comparatively higher melting temperatures (Nielsen et al., 1991). These properties and its similarities with DNA have led to PNA being proposed as a possible proto-RNA (Nielsen, 2007) in origin of life scenarios (Nelson et al., 2000; Ura et al., 2009). But which nucleic acid arose first and how? To answer these questions we must consider the conditions present on the early prebiotic Earth (Martin et al., 2008).

Assuming that the initial prebiotic synthesis of proto-biomolecules took place in aqueous solution and arose within the oceans of the Archean Earth (3.5 Ga ago), it is

likely that the chemical and temperature gradients at early hydrothermal vents may have provided the necessary driving forces for chemical reactions to occur (Martin and Russell, 2003). The production of simple organic molecules has been observed in present day vents (McCollom and Shock, 1997) and replicated in laboratory studies (Fu, 2007). For organic macromolecules to have been generated, small abiotic substances had to be concentrated to the extent that reactive centres came within sufficient proximity for further chemistry to have occurred, while the products would have needed to be protected from degradation. One way to realize these conditions is at the surfaces and within the internal pores of minerals present at such vent systems (Hazen and Sverjensky, 2010). A recent theory, by Hansma, suggests that layered clay minerals may have acted as the first proto-cells (Hansma, 2010).

Layered double hydroxides (LDHs) are a class of layered mineral, sometimes known as anionic clays (the sheets are positively charged and take up anions), which have been the focus of origins of life studies for a number of years (Arrhenius, 1993; Pitsch et al., 1995; Krishnamurthy et al., 1996; Arrhenius et al., 1997). It is thought that in the reducing conditions present on the early Earth, green rust ( $\text{Fe}^{2+}/\text{Fe}^{3+}$ ) LDHs may have been prevalent (Arrhenius, 2003). LDH minerals have structures that are conceptually derived from the layered mineral brucite ( $\text{Mg}(\text{OH})_2$ ), which has divalent  $\text{Mg}^{2+}$  metal cations octahedrally coordinated with hydroxide ions. In LDHs, the isomorphic substitution of trivalent in place of divalent cations causes a net positive charge which is counterbalanced by the introduction of anions within the interlayer region. Anions of particular interest to organic molecular evolution include  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{CO}_3^{2-}$ ,  $\text{PO}_4^{3-}$  and linear oligophosphates (Arrhenius et al., 1997). In the LDH interlayer, anions can be readily exchanged with other anionic species so that carbonates, cyanide, oligomeric phosphates and amino acid guests may all be intercalated into the LDH host structure.

Glyco-aldehyde phosphate (GAP) has been postulated as a candidate for a primitive building block *en route* to nucleic acids (Bean et al., 2006). These molecules become highly concentrated through intercalation into LDHs from a dilute external solution (Pitsch et al., 1995). Experimental results by Pitsch et al. (1995) show that LDHs are able to

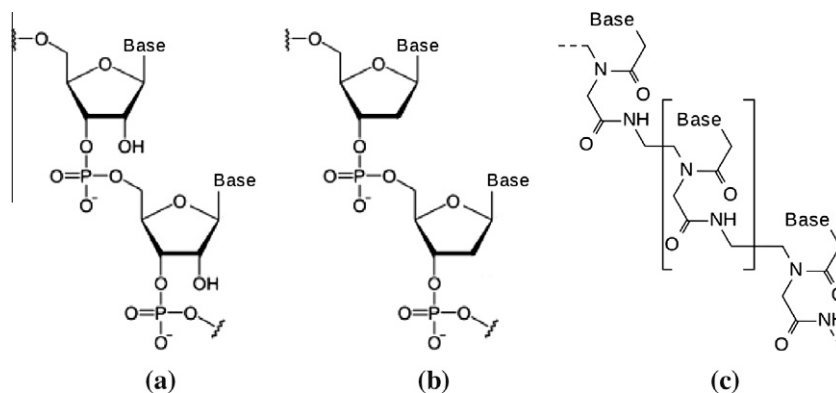


Fig. 1. The structural formulae of (a) ribonucleic acid (RNA), (b) deoxyribonucleic acid (DNA) and (c) peptide nucleic acid (PNA). PNA is an artificially synthesized nucleic acid. PNA has an uncharged backbone which is composed of repeating *N*-(2-aminoethyl)-glycine units linked by peptide bonds.

adsorb GAP anions from concentrations of a few  $\mu\text{M}$  to yield a  $\sim 10\text{ M}$  confined environment within the LDH interlayer. Subsequent condensation reactions of GAP ions in LDHs give a high yield of sugar phosphates (Krishnamurthy et al., 1996).

Owing to interest in LDH–DNA compounds for drug delivery (Choy et al., 1999, 2000, 2001, 2007; Kwak et al., 2002, 2004) and molecular code systems (Choy et al., 2004), several experimental studies have been undertaken reporting intercalation of DNA into LDHs. Experimental studies in these areas have provided some understanding of DNA–LDH systems; powder X-ray diffraction (PXRD) has demonstrated the changes in interlayer spacing in these materials (Desigaux et al., 2006), while circular dichroism and infrared spectroscopy have been used to study the intercalated nucleic acids (Oh et al., 2006). Although all forms of polyanionic DNA have been found to intercalate into LDHs (Desigaux et al., 2006), to the best of our knowledge no evidence of RNA or PNA intercalation into LDHs yet exists in the literature. Greenwell and Coveney further pointed out that, conceptually, LDHs might represent idealized information storage and transfer systems and that the separation distance of intralayer charge sites is remarkably similar to the distance between phosphate groups in nucleic acid structures (Greenwell and Coveney, 2006).

Origins of life studies have hitherto rarely used computer simulation techniques to understand the possible chemical pathways to the formation of the first biomolecules. However, computational methods provide powerful molecular level insights into the structure and properties of cationic and anionic clay based systems which are difficult to characterize experimentally (Greenwell et al., 2006). As a result, interest has begun to grow in the use of computational modelling to investigate systems of possible relevance to the origins of life (Thyveetil et al., 2008b,a; Mathew and Luthey-Schulten, 2010; Swadling et al., 2010). Furthermore, computer simulation can rapidly access environmental conditions difficult to attain by experiment.

In this study we use large-scale molecular dynamics simulation techniques, similar to those previously performed by us (Thyveetil et al., 2008b,a; Swadling et al., 2010), in order to compare the relative structural stability of DNA, RNA and PNA all within layered double hydroxides and free in bulk aqueous solution. We select a double stranded RNA to investigate the potential for prebiotic LDH–RNA systems to protect early forms of these nucleotides from extreme environmental conditions. We compare these results to those for double stranded DNA and PNA structures thus providing insight into the comparative properties of these nucleic acids in mineral and bulk aqueous environments.

## 2. METHODS

This section discusses the approach we employ to simulate the LDH–nucleic acid–water models. We explain the choice of forcefield needed to describe the interaction between the nucleic acids and LDH. We use large-scale molecular dynamics since the models we simulate contain up to around a quarter of a million atoms. A range of powerful

supercomputing resources were invoked to perform the simulations in as short a wallclock time as possible. We briefly describe the infrastructure and computational resources utilized in the [Electronic Annex](#).

### 2.1. Model construction

The layered double hydroxide which forms the basis for the mineral models employed in this study has unit formula  $[\text{Mg}_2\text{Al}(\text{OH})_6]_n\text{H}_2\text{O}\cdot\text{Cl}$  and is the same as that used in our recent studies (Thyveetil et al., 2008b,a, 2007). Chloride ions are present to counter the positive LDH charge. The LDH models were replicated from a unit cell with dimensions  $16.34\text{ \AA} \times 18.82\text{ \AA} \times 25.34\text{ \AA}$ , which was obtained by the refinement of powder X-ray diffraction data on hydroxalcite using Rietveld methods (Bellotto et al., 1996). In this work, models containing either 3 or 6 interlayer regions were simulated.

The sizes and sequences of the nucleic acid molecules used in this study are given in [Table 1](#). Three different forms of double stranded nucleic acid were constructed, as shown in [Fig. 2](#). The first structure was that of a dodecamer duplex of A-DNA assembled using the Nucleic Acid Builder, which is part of AmberTools (Case et al., 2011), with base-pair sequence  $d(5'\text{-CTTTTGCAAAAAG-}3')$  (Jha et al., 2005). The DNA sequence was chosen as it has been well characterized in past molecular dynamics simulations (Jha et al., 2005; Thyveetil et al., 2008b). The second structure was a dodecamer of A-RNA also built using the Nucleic Acid Builder with corresponding base-pair sequence  $d(5'\text{-CUUUUGCAAAAAG-}3')$  (the RNA sequence corresponds to the DNA sequence with thymine groups substituted by uracil). All phosphate groups have an unprotonated hydroxyl group and carry a  $-1$  negative charge, as we are assuming a pH of  $>7$  where all phosphate groups are known to be ionized. The third structure was a dodecamer of PNA, in the P-form, with the same base-pair sequence as the DNA dodecamer. The PNA strand's peptide backbone was terminated with N-terminus and C-terminus end-groups, as reported by Shields et al. (1998) and Soliva et al. (2000) in their MD studies of PNA:DNA and PNA:RNA molecules.

Nucleic acids RNA and DNA can adopt two different right-handed double helix structures, A and B. The A-form has a shorter more compact helical structure which has an increase in the number of base-pairs per rotation over the B-form. We chose the A-form of DNA to allow a direct comparison with the A-form used for RNA. The A-form of DNA is known to undergo a transition to the B-form (or something closely approximating it) in bulk water (see validation of nucleic acid models in the [Electronic Annex](#)). The A-form is known to be the preferred structure for DNA under conditions of high salinity and low hydration. The P-form of PNA, which is the name given to its native structure in aqueous solution (shown in [2](#)), appears structurally different to the A-form of DNA and RNA in [Fig. 2](#) because of the difference in composition of the peptide backbone in PNA from the phosphate backbone in DNA and RNA. The P-form can be characterized by the wide helix ( $28\text{ \AA}$ ) with almost twice the pitch ( $18\text{ bp}$  per

Table 1

Simulation cell compositions and dimensions for the Models I–VI containing MgAl–LDH intercalated with 12 bp linear duplex RNA, DNA or PNA, together with chloride ions and water. The large clay Models IV–VI were built to remove any finite size effects, including the suppression of thermal undulations in the clay sheets. Models VII, VIII and IX are the nucleic acids solvated in water with counterbalancing sodium ions. Models I–III were studied at a range of water contents, while Models IV–IX were studied at a fixed level of hydration at a range of temperatures and pressures between 300 K, 1 atm and 500 K, 100 atm.

Model	No. of atoms	Nucleic acid	No. of duplexes	Base sequences	Starting supercell dimensions (Å)	No. LDH sheets	Simulation time (ns)
I	12,650	RNA	1	d(5'-CUUUUGCAAAAAG-3') <sub>2</sub>	49.0 × 56.5 × 56.5	3	25
II	12,650	DNA	1	d(5'-CTTTTGCAAAAAG-3') <sub>2</sub>	49.0 × 56.5 × 56.5	3	25
III	12,650	PNA	1	d(5'-CTTTTGCAAAAAG-3') <sub>2</sub>	49.0 × 56.5 × 56.5	3	25
IV	249,600	RNA	24	d(5'-CUUUUGCAAAAAG-3') <sub>2</sub>	160.2 × 340.0 × 129.9	6	30
V	249,600	DNA	24	d(5'-CTTTTGCAAAAAG-3') <sub>2</sub>	160.2 × 340.0 × 129.9	6	30
VI	250,800	PNA	24	d(5'-CTTTTGCAAAAAG-3') <sub>2</sub>	160.2 × 340.0 × 129.9	6	30
VII	218,863	RNA	24	d(5'-CUUUUGCAAAAAG-3') <sub>2</sub>	124.3 × 126.7 × 145.0	–	30
VIII	221,101	DNA	24	d(5'-CTTTTGCAAAAAG-3') <sub>2</sub>	124.1 × 129.5 × 142.9	–	30
IX	230,159	PNA	24	d(5'-CTTTTGCAAAAAG-3') <sub>2</sub>	125.5 × 128.0 × 148.9	–	30

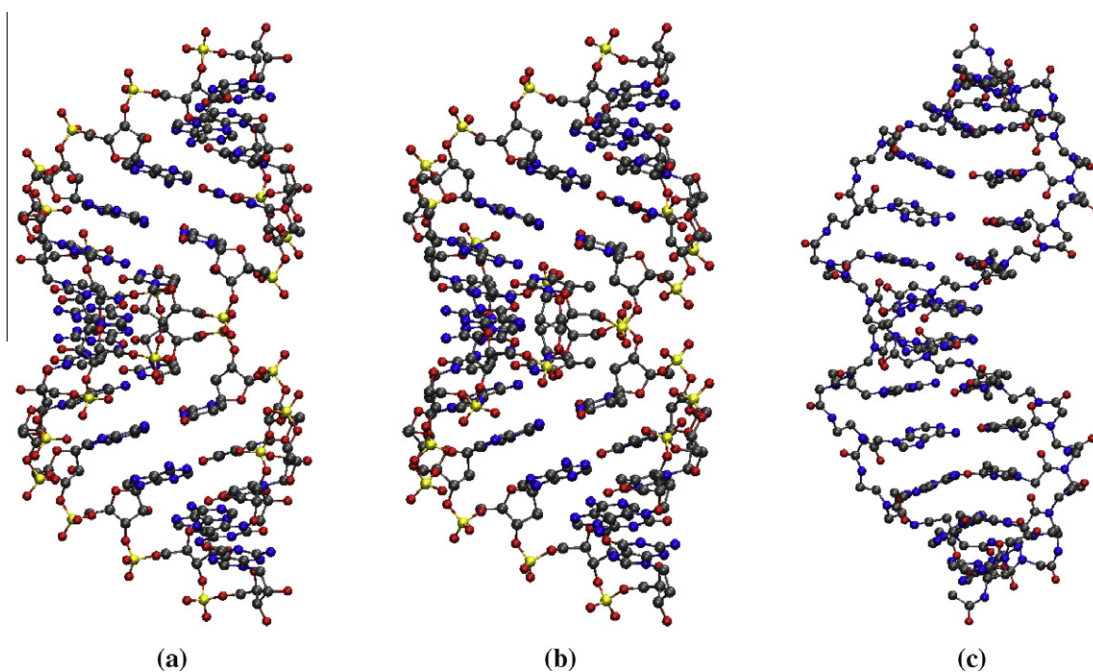


Fig. 2. A schematic three dimensional ball-and-stick representation of (a) A-RNA in Models I, IV and VII, (b) A-DNA in Models II, V and VIII and (c) PNA in Models III, VI and IX. All models are described in more detail in Table 1. Carbon, nitrogen, hydrogen, oxygen and phosphorus atoms are represented as grey, blue, red and yellow spheres, respectively. The structure of RNA differs from the structure of DNA by the additional hydroxyl group at the C2 position on the nucleotide ribose. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

turn) of an A- or B-form helix (10–11 bp per turn) (Nielsen et al., 1999). In order to make comparisons between the nucleic acids in water and when intercalated, simulations of the three types of nucleic acid were also performed solvated in water (Models VII–IX, see Table 1); the nucleic acids within these bulk water simulations were charge-balanced by sodium ions.

Many studies have shown that water plays a vital role in the structure of nucleic acids in aqueous solution; the role of water also governs the structural stability of nucleic acids once intercalated into layered double hydroxides (Choy et al., 2000; Desigaux et al., 2006; Thyveetil et al., 2008b).

The first part of our study therefore aims to investigate the impact of hydration on Models I, II and III, which consist of a nucleic acid intercalated within a LDH as shown in Fig. 3, by varying their water content. Models of varying degrees of hydration were created by changing the water content only in the interlayer containing the nucleic acid, while keeping the other interlayer at two water molecules per unit formula  $Mg_2Al(OH)_6$ . Previous studies of MgAl–LDH produced simulated hydration curves in general agreement with experimental data using the ClayFF force-field when intercalated with chloride ions (Wang et al., 2001; Thyveetil et al., 2007), and with DNA (Thyveetil

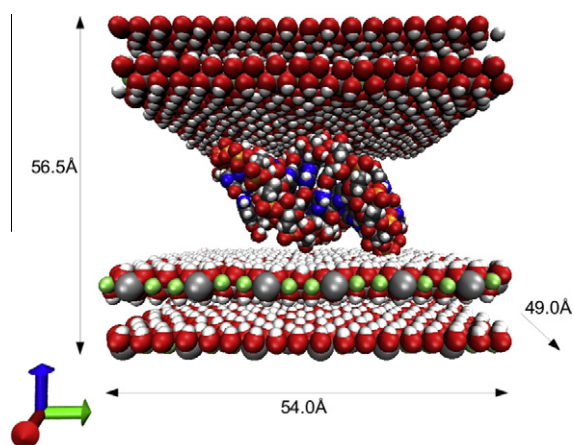


Fig. 3. Initial structure of the LDH–RNA Model I at the start of the simulation. Model I consists of three  $Mg_2Al$  LDH sheets with a 12 bp RNA duplex placed within one of the interlayers. For clarity, water molecules and chloride ions are not displayed. Magnesium, aluminium, oxygen, hydrogen, nitrogen, carbon and phosphorus are displayed as green, silver, red, white, blue, grey and orange spheres respectively. The  $x$ ,  $y$  and  $z$  axes are displayed as red, green and blue arrows, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

et al., 2008b). In order to create hydration curves for the RNA, DNA and PNA models, 16 different LDH–nucleic acid models were built with an increasing number of water molecules, from two water molecules per Al atom up to 16 water molecules per Al atom. In cases where nucleic acids are included in adjacent interlayers, the molecules were offset with respect to each other. Our previous simulations of DNA–LDH have shown that the model adopts the lowest potential energy when the nucleic acids in adjacent interlayers are not stacked directly on top of each other (Thyveetil et al., 2008b,a).

We have shown that small LDH models suffer from finite size effects which cause the suppression of thermal undulations. The collective thermal motion of atoms in the LDH sheets causes undulatory modes. The largest wavelength of thermal undulations was found to be 40 Å (Thyveetil et al., 2007). The lateral dimensions of the largest models used in this present study were more than twice this distance. Model IV contained an RNA dodecamer duplex while Model VI contained a PNA duplex of similar length; we sought the hydration level which yielded basal spacings comparable with experimental findings for analogous DNA–LDH models, which are 21.1–23.9 Å. Experimental results were taken from the variation in the  $d$ -spacing as a function of the number of water molecules calculated from the thermogravimetric analyses recorded in a static air atmosphere with a heating rate of 278 K/min, in the temperature range of 298–1073 K (Choy et al., 2000; Desigaux et al., 2006). Larger models based on this experimental hydration state were created containing six LDH interlayers for the RNA, DNA and PNA duplex models (Models IV–VI, Fig. 4), with four RNA/DNA/PNA duplces in each interlayer, in order to simulate the nucleic acid–LDH models at higher temperatures and pressures. Models IV–IX

were simulated at five different conditions: 300 K at 1 atm, 350 K at 1 atm, 400 K at 50 atm, 450 K at 100 atm and 500 K at 100 atm, in order to study the stability of nucleic acids intercalated and free in bulk water under different conditions associated with possible origins of life environments (Thyveetil et al., 2008b). The various temperature and pressure conditions were selected to span mild and extreme conditions associated with hydrothermal vent models and mineral mediated origins (Martin et al., 2008).

## 2.2. Potential parameterization

The potential energy of the model is parameterized using a hybrid approach that combines the ClayFF (Cygan et al., 2004), a forcefield for the simulation of minerals such as LDHs, and the Amber parm99 forcefield (Wang et al., 2000) widely used to simulate DNA and RNA (Hobza et al., 1997; Young et al., 1997). In order to simulate PNA, we utilized the forcefield developed by Shields et al. (1998) which was found to reproduce PNA structural properties with good accuracy. The nucleic acids and LDH interact via non-bonded forces only. The missing Lennard–Jones intermolecular potential parameters are supplemented using Lorentz–Berthelot mixing rules, which suffice for models dominated by electrostatics (Cormack et al., 2004; Duffy and Harding, 2004; Harding and Duffy, 2006), as shown in our earlier studies (Suter et al., 2007; Thyveetil et al., 2008b,a; Swadling et al., 2010). Water molecules are described using the flexible single point charge (SPC) model (Berendsen et al., 1981). Because the ClayFF and Amber forcefields are parameterized at ambient temperatures and pressures, the forcefields are likely to be less reliable at the higher temperatures and pressures reported here. Nevertheless, the higher temperature and pressure simulations provide important qualitative comparisons between the different nucleic acid models and furnish insight into the effects of temperature and pressure on these models, as previously reported (Thyveetil et al., 2008b). A discussion on the reliability of nucleic acid forcefields under various conditions (as well as at long timescales) is given by Cheatham and Young (2001).

## 2.3. Molecular dynamics

The models were simulated using the large-scale atomistic/molecular massively parallel simulator (LAMMPS) (Plimpton, 1995) a highly scalable, parallel molecular dynamics program which can be run on a variety of computer architectures. We energy-minimized the models using the steepest descent method before carrying out the MD simulations using the NPT ensemble (which means the number of particles ( $N$ ), the pressure ( $P$ ) and the temperature ( $T$ ) are kept at a constant value), using techniques discussed in our previous studies (Thyveetil et al., 2007, 2008b; Swadling et al., 2010). After equilibration was deemed to have been established, production runs were performed for 30 ns to provide enough statistics to capture diffusion data for the nucleic acid molecules (Thyveetil et al., 2007). The models were deemed to have reached equilibration before 500 ps by monitoring the model potential en-

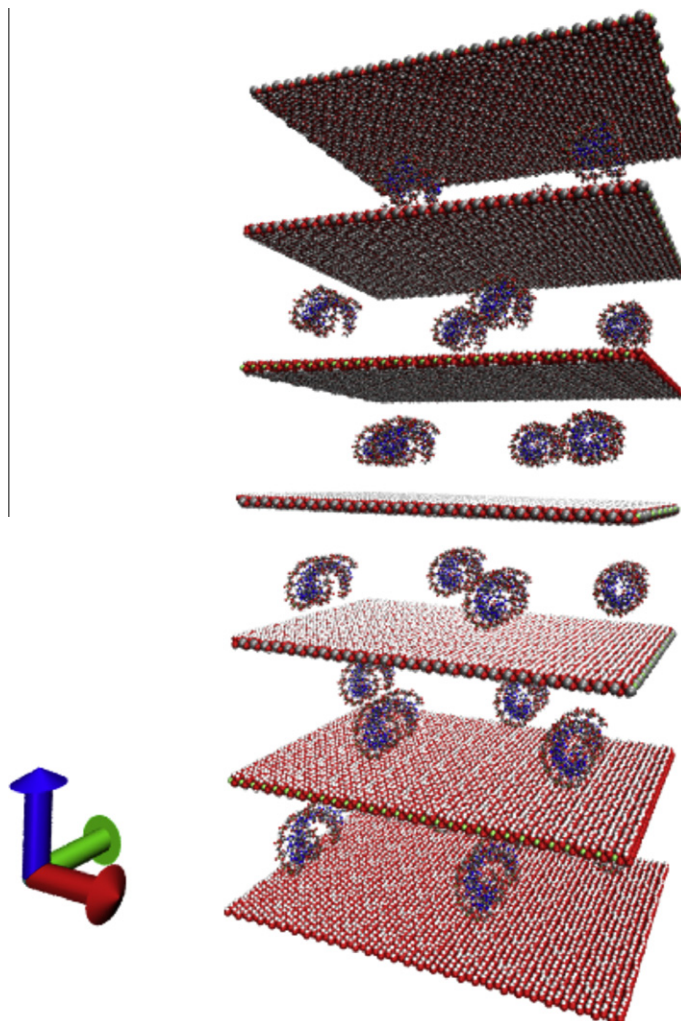


Fig. 4. Initial structure of the large LDH–RNA Model IV at the start of the simulation. The model consists of  $6\text{Mg}_2\text{Al}$  interlayers with four 12 bp RNA double strands per interlayer. See Table 1 for more details of sequence and structure. For clarity, water molecules and chloride ions are not displayed. Colour scheme is the same as that used in Fig. 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

ergy and the cell parameters (see Figs. 2–4 in the [Electronic Annex](#)). The data show that all models are at equilibrium (or at least in long lasting steady states) as adjudged by the figures of merit, including potential energy and cell parameters, and that any further simulation beyond 30 ns is not necessary. A Nosé–Hoover thermostat/barostat was used to regulate the temperature and pressure of each simulation.

#### 2.4. Analysis of simulations

Post processing analysis of the accumulated MD trajectory data was performed using the 3DNA software analysis tool (Lu and Olson, 2003) to quantify the overall changes in the structure, the root mean square deviation (RMSD) being calculated by removing translational and rotational degrees of freedom through a least-squares fit followed by rotation to the original reference frame of the molecule, using the relation:

$$r_{\text{RMSD}}(\mathbf{r}, \mathbf{r}_0) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_{i,0})^2}, \quad (1)$$

where  $\mathbf{r}$  is the current position of atom  $i$ , and  $\mathbf{r}_{i,0}$  is its initial position. Initial atomic coordinates were taken from the nucleic acid builder-generated molecular structures for the RNA, DNA and PNA duplex models, and the summation carried out over all  $N$  nucleotide atoms.

The radius of gyration,  $r_G$ , estimates the molecule's effective size during the simulation:

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

where  $\mathbf{r}$  is the current position of atom  $i$ , and  $\langle \mathbf{r}_i \rangle$  is the mean position of the atom.

Principal component analysis (PCA) was used to describe the main changes in nucleic acid structure with time by identifying the dominant collective modes of motion (Amadei et al., 1993; Hess, 2000; de Groot et al., 2001;

Jha et al., 2005). In a Cartesian coordinate system, the covariance matrix can be defined as:

$$\mathbf{C} = \langle (\mathbf{r} - \langle \mathbf{r} \rangle)(\mathbf{r} - \langle \mathbf{r} \rangle)^T \rangle, \quad (3)$$

where  $\mathbf{r}$  represents the atomic positions of the nucleic acid 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. The dominant modes of motion for DNA in water are, as shown in our previous publication (Thyveetil et al., 2007), twisting of the helix around a central axis, junction bending and wedge bending.

Detailed structural information regarding the Watson–Crick hydrogen bonds in the three nucleotide models was gleaned using the 3DNA software analysis tool (Lu and Olson, 2003) by least-squares comparison with crystal structures of model nucleic acids. In order to identify a Watson–Crick base pair, the following geometric criteria must be met: the distance between the origins of two bases must be less than 15 Å, the vertical separation must be less than 2.5 Å, the normal vectors of the base-pairs must be rotated less than 65.0° with respect to each other and there must be at least one pair of nitrogen/oxygen base atoms that are within 4.05 Å of each other (Lu and Olson, 2007). The interlayer structure was also analyzed using radial distribution functions as described in our previous work (Thyveetil et al., 2007, 2008b). The radial distribution function describes how the atomic density varies as a function of the distance from a central atom.

In order to compare the mobility of the nucleic acid strands within the LDHs, the self-diffusion coefficients of all nucleic acid duplex strands were calculated for Models IV–VI, and averaged over the number of duplexes in each model. The diffusion coefficient is calculated from the slope of the mean square displacement ( $\delta r^2$ ) vs. time graph. The MSD was computed from the centre of mass of each DNA duplex strand in order to include only translational motion in the calculation. The graph was constructed using multiple time origins in order to improve the statistics of the calculation. As the DNA molecules are constrained to move within the interlayer planes the diffusion is confined to the  $xy$ -plane. Therefore, we compute values for the two dimensional diffusion coefficient, which is given by  $4D_{xy} = \frac{d}{dt} \delta r^2$  where  $\delta r^2 = \delta x^2 + \delta y^2/2$ . The slope of the  $\delta r^2$  vs. time plot exhibits a linear region of slope  $D$ , and we find the error on the least square fit to estimate the error of the diffusion coefficient.

### 3. RESULTS

As reported in our previous work on DNA–LDH models (Thyveetil et al., 2008b), the nucleic acids studied here do not exhibit large changes in position and conformation within the interlayer over the duration of the simulation; conversely the LDH sheets themselves show evidence of their flexibility around bulky intercalants (Thyveetil et al., 2008a,b). Visualization of the final structures of the larger

models (Models IV, V and VI) at ambient conditions are shown in Fig. 5.

Section 3.1 describes in detail the hydration behaviour of LDH models containing RNA, DNA and PNA, and compares these properties to our previous results for similar sized DNA–LDH models (Thyveetil et al., 2008b). In Section 3.2, we report the behaviour of RNA, DNA and PNA intercalated in larger LDH structures, which eliminate finite size effects and capture emergent thermal undulations within the LDH sheets (Thyveetil et al., 2007). Previous experimental work and simulations of DNA, RNA and PNA in bulk aqueous solution were used to validate the nucleic acid models reported here; the data is presented in the [Electronic Annex](#).

#### 3.1. Effect of hydration on nucleic acid–LDH models

The swelling/hydration curves were produced by simulating Models I–III with an increasing amount of water molecules per nucleic acid containing interlayer, beginning with a dehydrated model and adding two water molecules per LDH unit formula ( $[\text{Mg}_2\text{Al}(\text{OH})_2]$ ) up to 16 water molecules per LDH unit. The variable  $n$  refers to the number of water molecules per unit formula. Each hydrated model was simulated for 3 ns once equilibrium was deemed to have been established.

Computing the hydration curves (clay layer separation as a function of water content) gives insight into how the nucleic acid structures change with varying amounts of water. Fig. 6 shows the direct comparison between the change in basal spacing, RMSD from the crystal structure and disruption of Watson–Crick hydrogen bonded base-pairs, with increasing water content for DNA, RNA and PNA intercalated in LDH. Some similarity exists in basal spacings between DNA and PNA, whereas on average RNA has generally lower basal spacings, especially when the number of water molecules per unit formula ranges between  $n = 6$ –15.

At low hydration states, below 10 water molecules per Al atom, significant structural deformation can be seen as evidenced by the relatively high RMSD and correspondingly low percentage of Watson–Crick hydrogen bonds remaining intact, but as the water content of the interlayer reaches similar values to bulk water, at 16 water atoms per Al, the nucleic acids return to their bulk-water solvated double helix structures (see Fig. 7). Plots of RMSD vs. hydration in Fig. 6 show that the structures of both RNA and PNA become increasingly similar to that of the same nucleic acid in bulk water as the interlayer water content increases. However, the RMSD values of RNA and PNA are significantly greater than DNA after  $n = 8$ . In addition, the RMSD between the intercalated and reference structures does not decay monotonically, but rather exhibits numerous local minima and maxima. These “steps” are also observed in Fig. 6a for the basal spacing and are attributed to hydrogen bonding networks formed between water and the clay surfaces, which only allow expansion of the interlayer once distinct hydration layers are fully filled (Boek et al., 1995b,a; Wang et al., 2001; Thyveetil et al., 2007). This “step” behaviour is prominently seen in Fig. 6c, where

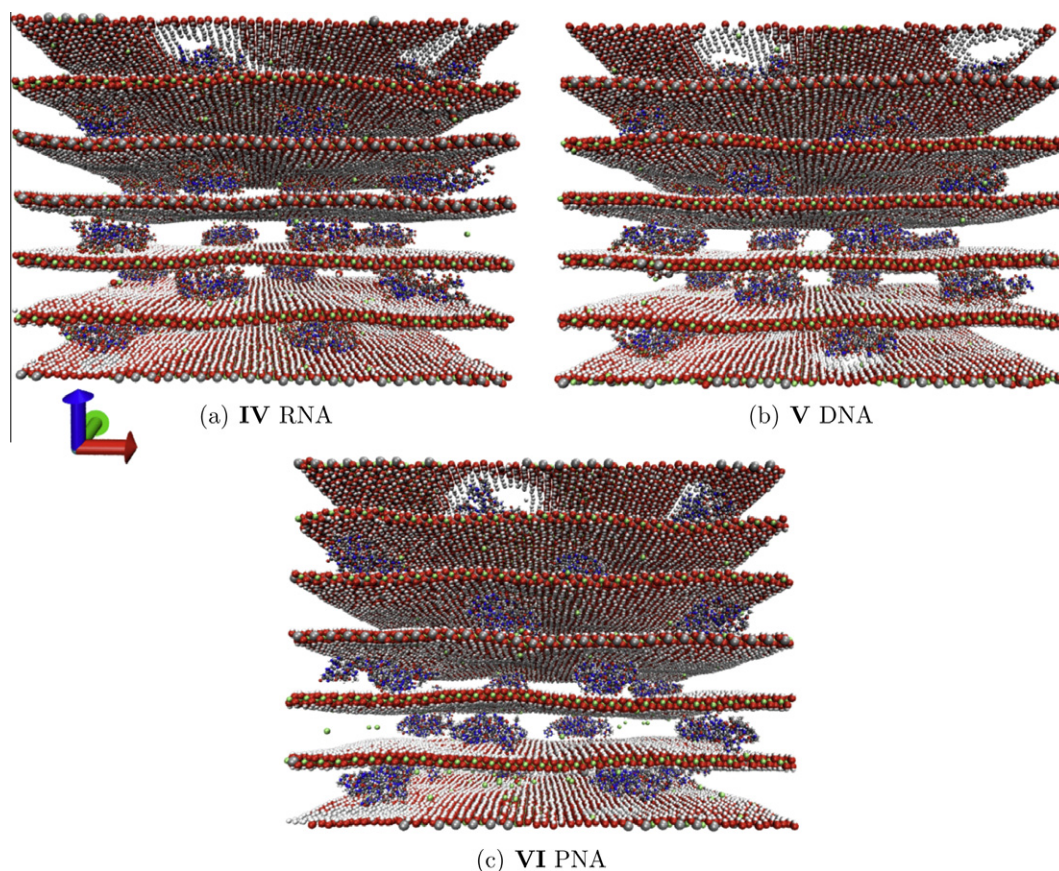


Fig. 5. Final structures of the large LDH–nucleic acid models: (a) Model IV, (b) Model V and (c) Model VI after the 30 ns production phase of each simulation. The colouring scheme used is the same as that of Fig. 3, with water and chloride ions not displayed to aid viewing. All nucleotide motion within the LDH sheets is significantly restricted compared to that in bulk water. In addition, visualization reveals properties such as thermal undulations in the LDH sheet, as well as corrugation of the sheets around the nucleotides. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

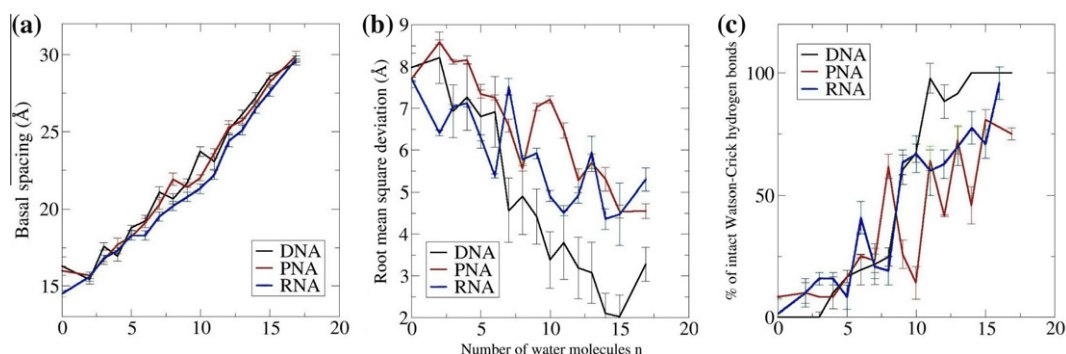


Fig. 6. Comparison of the hydration properties of intercalated RNA duplex in Model I, DNA in Model II and PNA in Model III, at 300 K and 1 atm: (a) interlayer (basal) spacing, (b) root mean square deviation (RMSD) relative to the nucleic acid crystal structure and (c) percentage of Watson–Crick bonded base-pairs remaining. Error bars show the standard deviation for each value. From the change in basal spacing in (a), it appears that the intercalated RNA double strands are the most easily compressed, possibly owing to the different base-pair sequence resulting in weaker hydrogen-bonding between the strands compared to those of the DNA and PNA models. However, in general the intercalated PNA duplex strands have the largest RMSD values (b), indicating that the base-pairing in this model is significantly disrupted, with decreasing interlayer water, compared to the DNA and RNA duplexes, as confirmed by (c).

the percentage of intact Watson–Crick hydrogen bonds is reported, and is particularly obvious for PNA. From the results for RMSD values, it can be inferred that intercalated

PNA strands show lower structural stability, in terms of number of Watson–Crick base-pairs, at higher hydration states ( $n > 8$ ) compared to RNA and DNA.



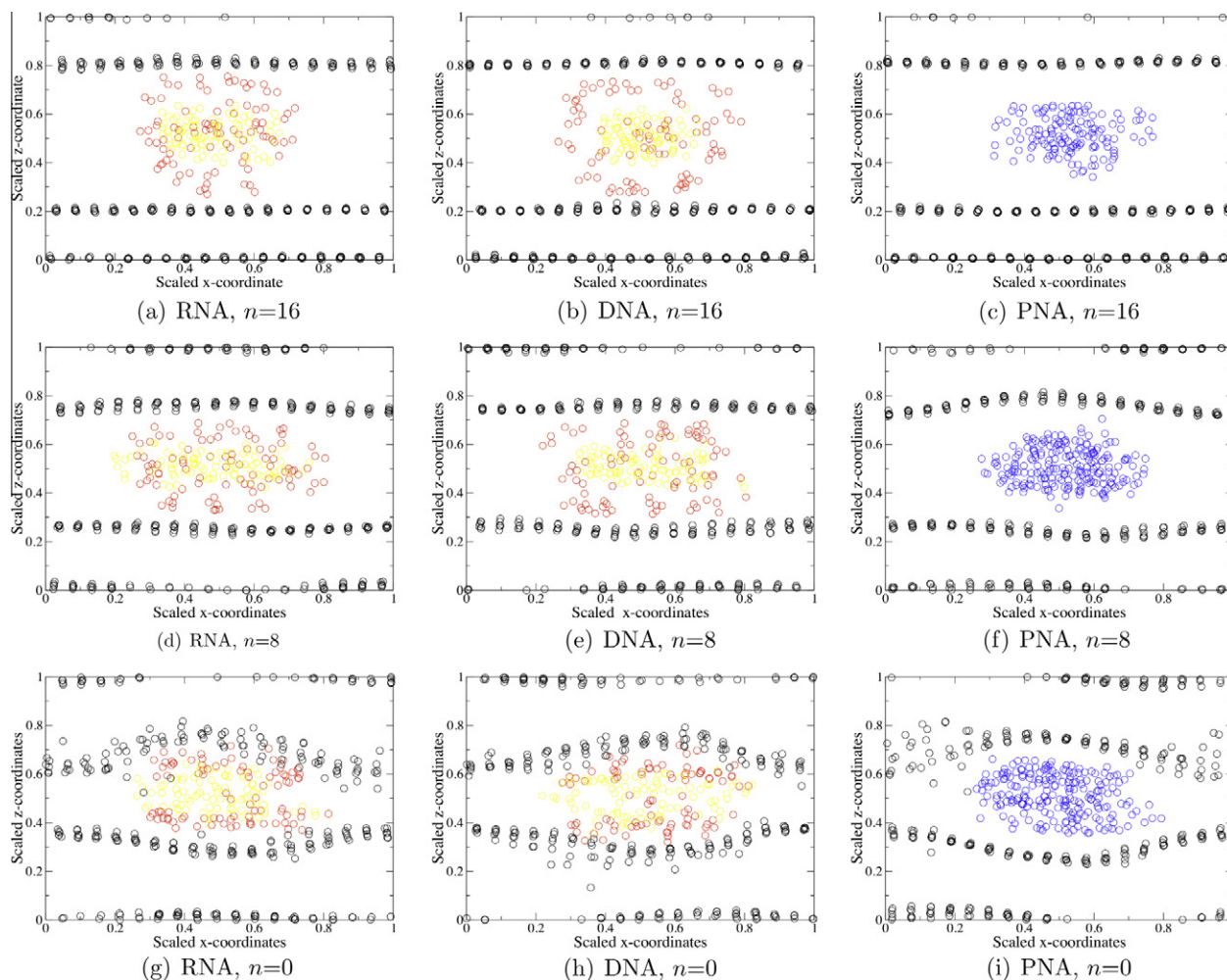


Fig. 7. Cross-section in the  $xz$  plane of Models I–III for various hydration states at 300 K and 1 atm, taken from the final snapshot of the simulation. Black circles correspond to positions of aluminium ions in the LDH sheets. Periodic boundaries cause some aluminium ions to reside at the top of the simulation cell. Red, yellow and blue circles correspond to phosphate groups, nitrogen atoms in the ribose sugar and the nitrogen atoms belonging to the PNA peptide backbone respectively. The variable  $n$  refers to the number of water molecules per unit formula  $[\text{Mg}_2\text{Al}(\text{OH})_6]_n\text{H}_2\text{O}$ . (a–c) RNA, DNA and PNA, respectively, hydrated with  $n=16$ ; (d), (e and f) RNA, DNA and PNA, respectively, hydrated with  $n=8$  and (g–i) dehydrated RNA, DNA and PNA, respectively. The cross-section of the full hydrated PNA model, which is shown in subfigure (c), shows the PNA adopting a different conformation to the circular tube-like structure of RNA and DNA. This observed difference may be due to  $\text{Cl}^-$  ions assembling near to the LDH sheet surface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

Presently, there are no experimental results which describe the basal spacing of RNA–LDH or PNA–LDH models as a function of water content; instead, the results from our DNA–LDH models are used to give a rough estimate as to which hydration state to use (Choy et al., 1999; Choy and Son, 2004; Desigaux et al., 2006), when building larger simulation models. The experimental value of the basal spacing under ambient conditions for DNA–LDH models is reported to be between 21.1 and 23.9 Å (Choy et al., 2000; Desigaux et al., 2006), which corresponds to a water content,  $n$ , between 10 and 13 water molecules per unit formula for RNA–LDH models and 8 and 11 for PNA–LDH. As RNA–LDH models have lower basal spacings compared to PNA–LDH and DNA–LDH models (Fig. 6), these LDH–nucleic acid models were built with  $n=10$  water molecules per unit formula  $[\text{Mg}_2\text{Al}(\text{OH})_6]_n\text{H}_2\text{O}$ . As the

PNA double strand is less easily compressed, the number of water molecules per unit formula chosen was  $n=9$ , to ensure values are comparable with experimental basal spacings for DNA–LDH models. The following sections describe the results obtained for larger LDH models (Models IV, V and VI), each in turn containing a different nucleic acid moiety, at different temperatures and pressures.

### 3.2. RNA, DNA and PNA intercalants

Model IV was used to study the properties of RNA strands intercalated within a six layer LDH supercell, with four double strands per interlayer. Under ambient conditions (300 K and 1 atm) Table 2 shows that, on average, base-pairing is substantially less than for RNA in bulk water, indicating that the environment within the LDH

Table 2

Comparison of structural parameters for RNA, DNA and PNA duplexes intercalated in LDH and free in aqueous solution (Models IV–IX). The mean RMSD is calculated by comparing, from the initial structure, the coordinates of the nucleic acid over 30 ns of MD with that of the original structure; the error is computed from the standard deviation. Of the three nucleic acids under investigation, DNA is observed to be the most stable when intercalated within an LDH, with respect to the percentage of Watson–Crick base-pairs. The opposite is found in bulk water where both PNA and RNA are shown to be more stable than DNA. These trends suggest that clay minerals could have played a major role in mediating the transition from RNA to DNA as the main informational biomolecule.

Model	Temperature (K)	Pressure (atm)	Mean RMSD (Å) per duplex	Mean % of intact Watson–Crick bonds per duplex	Maximum basal spacing (Å)
IV (RNA–LDH)	300	1	4.7 ± 0.9	26.3 ± 1	22.9 ± 0.03
	350	1	6.2 ± 0.3	21.2 ± 1	22.7 ± 0.03
	400	50	6.7 ± 0.5	17.1 ± 1	22.8 ± 0.03
	450	100	7.0 ± 0.5	13.7 ± 2	24.0 ± 0.07
	500	100	8.1 ± 0.8	5.1 ± 0.7	27.2 ± 0.7
VII (RNA in bulk water)	300	1	2.5 ± 0.5	89.7 ± 4	–
	350	1	6.9 ± 1.5	82.6 ± 7	–
	400	50	9.3 ± 1.8	68.2 ± 7	–
	450	100	13.3 ± 0.5	55.2 ± 5	–
	500	100	11.2 ± 1.0	14.7 ± 14	–
V (DNA–LDH)	300	1	4.3 ± 0.1	62.1 ± 3	19.2 ± 0.02
	350	1	4.3 ± 0.1	54.8 ± 6	20.3 ± 0.02
	400	50	4.4 ± 0.1	57.8 ± 4	19.6 ± 0.02
	450	100	4.4 ± 0.1	39.0 ± 2	21.0 ± 0.02
	500	100	4.4 ± 0.1	18.5 ± 8	22.2 ± 0.03
VIII (DNA in bulk water)	300	1	4.8 ± 0.3	82.0 ± 1.0	–
	350	1	5.5 ± 0.4	74.1 ± 1.9	–
	400	50	4.8 ± 0.5	40.5 ± 1.9	–
	450	100	5.7 ± 0.2	25.1 ± 2.5	–
	500	100	4.9 ± 1.0	14.7 ± 3.3	–
VI (PNA–LDH)	300	1	6.9 ± 0.2	23.3 ± 2.1	23.0 ± 0.03
	350	1	7.4 ± 0.2	26.1 ± 2.2	23.9 ± 0.3
	400	50	7.6 ± 0.2	28.0 ± 2.1	25.0 ± 0.04
	450	100	7.9 ± 0.2	19.7 ± 1.5	25.0 ± 0.05
	500	100	9.2 ± 0.3	14.4 ± 3.9	27.7 ± 0.10
IX (PNA in bulk water)	300	1	1.5 ± 0.4	99.9 ± 0.4	–
	350	1	6.3 ± 3.0	99.9 ± 1.1	–
	400	50	2.7 ± 0.5	99.8 ± 1.2	–
	450	100	9.9 ± 1.7	49.2 ± 1.1	–
	500	100	10.5 ± 1.3	14.9 ± 4.3	–

interlayers causes the Watson–Crick hydrogen bonding to be significantly disrupted. Table 2 provides structural information for a DNA–LDH model; the intercalated DNA strands retain ~60% of all base-pairing up to 450 K and 100 atm, while that for RNA strands decreases monotonically from a maximum of 25% with increasing temperatures and pressures. RNA within bulk water appears to have better retention of Watson–Crick base-pairing than DNA and PNA in bulk water, but the RMSD values are greater above 400 K and 50 atm than for the intercalated RNA, indicating that the LDH sheets constrain the structure of intercalated RNA molecules at elevated temperature and pressure, albeit at the expense of Watson–Crick pairing.

In Fig. 8a, we show the RMSD of RNA intercalated within the LDH, as a function of simulation time, relative to the A-form and B-form crystal structures of RNA. The RNA is found to deviate only by a few Å from the A-form crystal structure at the highest hydration level investigated. At all other lower levels of hydration the RNA structure

deviates much more (5–6 Å) from the A-form crystal structure, and even more so from the B-form.

In Fig. 8b, we show the RMSD of DNA intercalated within the LDH, as a function of simulation time, relative to the A-form and B-form crystal structures of RNA. Our findings indicate that the DNA in a fully hydrated LDH clay adopts a structure intermediate between those of the crystalline A- and B-forms. At lower levels of hydration, the confinement and the clay environment lead to a greater distortion of the nucleic acid; the equilibrated structures are closer to the A-form than the B-form.

In Fig. 8c, we show the RMSD of PNA whilst intercalated within an LDH, as a function of simulation time, from the crystal structure of PNA, which we refer to as the P-form, at various levels of hydrations. The intercalated PNA most resembles the initial P-form when simulated with 16 water molecules per unit formula (i.e. at higher hydration levels). At lower hydration levels the structure is considerably distorted from the P-form.

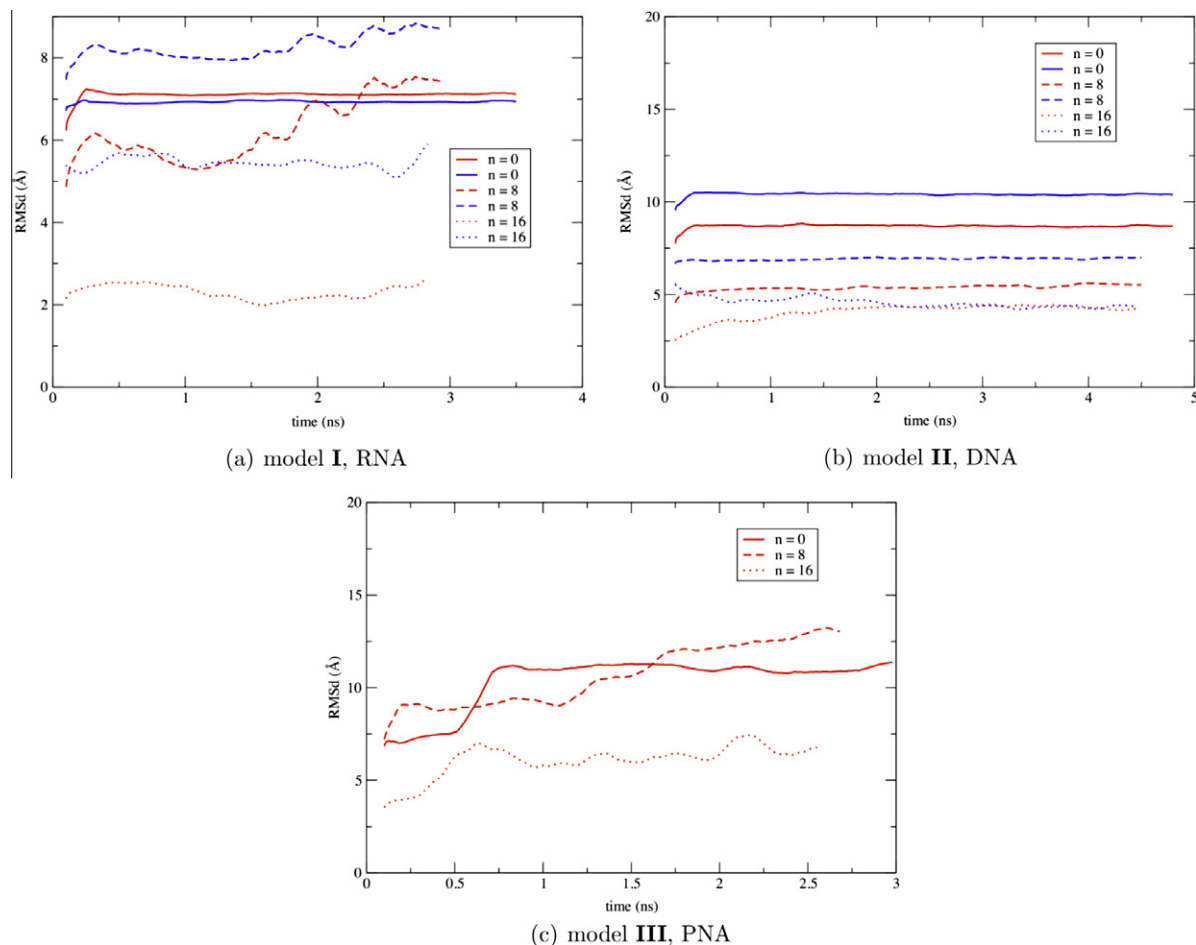


Fig. 8. Root mean squared deviation (RMSD) of nucleic acids, RNA and DNA, intercalated within the interlayer of an LDH, from their respective crystal structures in the A-form (shown in red) and in the B-form (shown in blue), at various levels of hydration. (c) The RMSD of PNA, intercalated within the interlayer of an LDH, from the initial P-form. The variable  $n$  refers to the number of water molecules per unit formula. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

The radii of gyration for all three duplex nucleotides intercalated in a LDH and in bulk water at ambient conditions, shown in Fig. 9, demonstrate that PNA has the largest relative size both when free in bulk water and intercalated within the LDH. There is a change in the relative order of size between the three different nucleic acids at ambient conditions when intercalated within the LDH compared to the nucleic acids in bulk water; DNA has the smallest relative size when intercalated (Model V) whilst RNA has the smallest size when free in bulk aqueous solution (Model VII). At elevated temperatures and pressures RNA, DNA and PNA within the LDH (Models IV–VI) take on the same order of size as at ambient conditions (see Fig. 6 in the *Electronic Annex*), whilst the nucleic acids free in bulk aqueous solution at elevated temperatures and pressures have a reversed order of size compared to the order at ambient conditions. At 450 K and 100 atm PNA has the smallest size in bulk water relative to RNA and DNA, while they all show increased fluctuations in size in bulk water compared to that under ambient conditions, indicating that the extreme conditions alter the size and structure of the nucleic acid in bulk water. The relative duplex nucleotide size shown by the radius of gyration at elevated

temperatures and pressures further indicates that the LDH protects the structure and conformation of the nucleic acid from changes caused by extreme environments.

The RDF plots in Fig. 10 provide insight into the overall structure of the RNA–LDH model and how it varies with increasing temperature. Under ambient conditions, the RDF plots show that, as expected, LDH surface hydrogen atoms are closest to phosphorus atoms in negatively charged phosphate groups (Thyveetil et al., 2008b). Although the peak intensity of the RDF curves for aluminium and magnesium ions are very similar, the intensities corresponding to the closest surfaces,  $r = 5.3 \text{ \AA}$  for  $\text{Al}^{3+}$  ions and  $r = 5.4 \text{ \AA}$  for  $\text{Mg}^{2+}$  ions, manifest a slight preference for proximity to aluminium compared with magnesium ions. At higher temperatures and pressures, this behaviour in the RDF plots is enhanced, suggesting greater interaction of the phosphorus and LDH atoms. Conversely, peak intensities for RDF plots of water surrounding phosphorus atoms diminish with increasing temperature and pressure.

RDF plots in Fig. 11, centred on the phosphorous atoms within the nucleic acids, except for PNA which is centred on the  $\text{sp}^2$  hybridized amide carbonyl carbon atom in the PNA

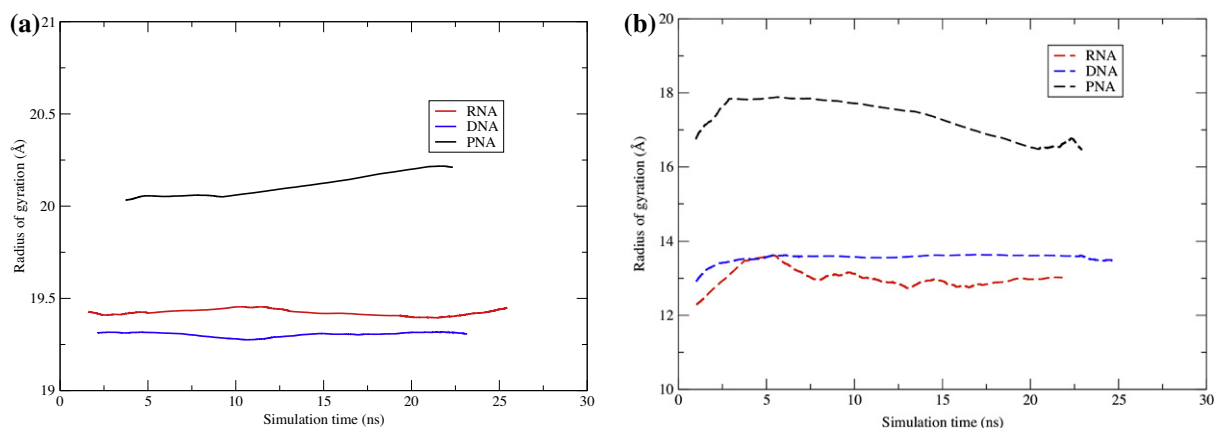


Fig. 9. Radii of gyration of nucleic acids intercalated within an LDH for (a) Model IV (solid red), Model V (solid blue) and VI (solid black) at 300 K and 1 atm. Radii of gyration of nucleic acids in bulk water for (b) Model VII (dashed red), Model VIII (dashed blue) and IX (dashed black) at 300 K and 1 atm. Compared to RNA and DNA, the relative size of PNA is the largest both when free in bulk water and intercalated within the LDH. There is a change in the relative order of size between the three different nucleic acids at ambient conditions when intercalated within the LDH compared to the nucleic acids in bulk water; DNA has the smallest relative size when intercalated whilst RNA has the smallest relative size when free in bulk aqueous solution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

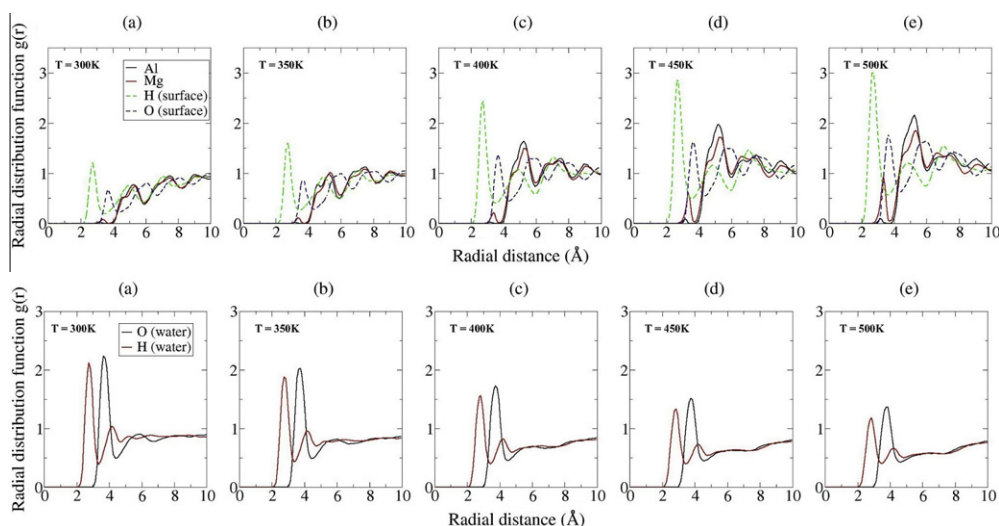


Fig. 10. Average radial distribution functions for Model IV, which consists of RNA intercalated LDH, between phosphorus atoms in phosphate groups and LDH atoms (top row) and water molecules (bottom row), at conditions of (a) 300 K and 1 atm, (b) 350 K and 1 atm, (c) 400 K and 50 atm, (d) 450 K and 100 atm and (e) 500 K and 100 atm. With increasing temperature, the peaks within the RDF for LDH atoms increase in amplitude, suggesting greater interaction of phosphate groups with the LDH surface. The RDF peak intensities decrease for water; presumably hydrogen bonding between phosphate groups and water molecules becomes increasingly disrupted at higher temperatures and pressures.

backbone (see Fig. 1) show the peak intensities for the RDF plots between phosphate group P atoms and selected LDH atoms increase with increasing temperature, suggesting that the phosphate groups in PNA have stronger interactions with the LDH surface at higher temperatures and pressures. By contrast, the peak intensities in the RDF plots for phosphate group P atoms and water O and H atoms decrease with increasing temperatures and pressures. This indicates that hydrogen bonding between phosphate groups and water molecules is disrupted with increasing temperature. This behaviour is also seen in Model IV. However, the principal component analysis (PCA) plots shown in Fig. 12

reveal a much more well defined structure and show little difference to the behaviour obtained for PNA in bulk water (see Fig. 13). This figure shows, *inter alia*, that the amplitude of motion along the first eigenvector is  $\sim 8$  orders of magnitude larger for PNA in bulk water compared to PNA intercalated within LDH, reflecting the heavily constrained motion of PNA inside LDH sheets (Fig. 12).

Fig. 15 shows that the main contribution to the dynamics for intercalated RNA and DNA arises from the motion of the phosphate groups furthest away from the LDH. These phosphate groups oscillate towards/away from the central axis, tensioned by opposing forces of attraction

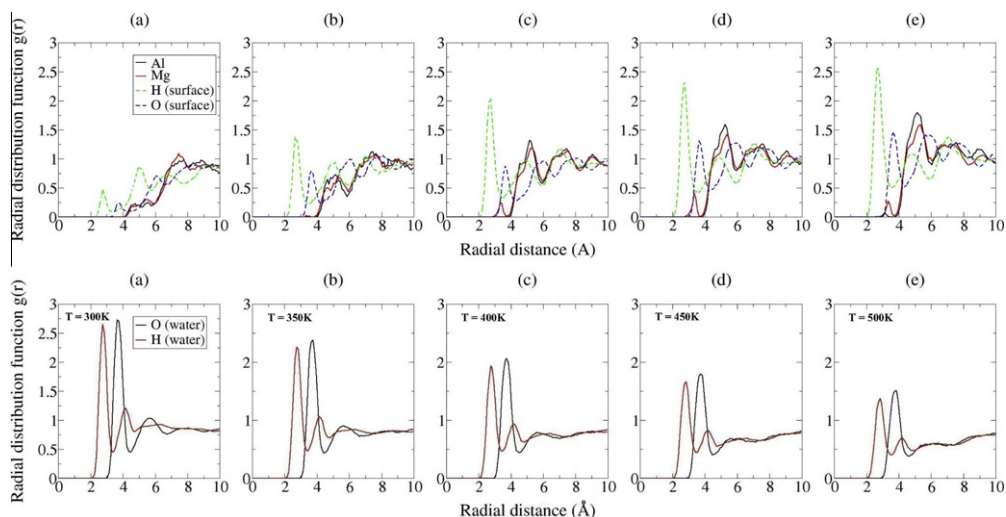


Fig. 11. Average radial distribution functions for specified atoms (indicated in legend at top right of first panel in each row) in Model VI, which consists of PNA intercalated LDH, at (a) 300 K and 1 atm (b) 350 K and 1 atm (c) 400 K and 50 atm (d) 450 K and 100 atm (e) 500 K and 100 atm, calculated with respect to the phosphorus atoms in nucleic acid phosphate groups. With increasing temperature, the peaks within the RDF for LDH atoms increase, suggesting greater interaction of phosphate groups with the LDH surface. The RDF peak intensities decrease for water; presumably hydrogen bonding between phosphate groups and water molecules becomes increasingly disrupted at higher temperatures and pressures.

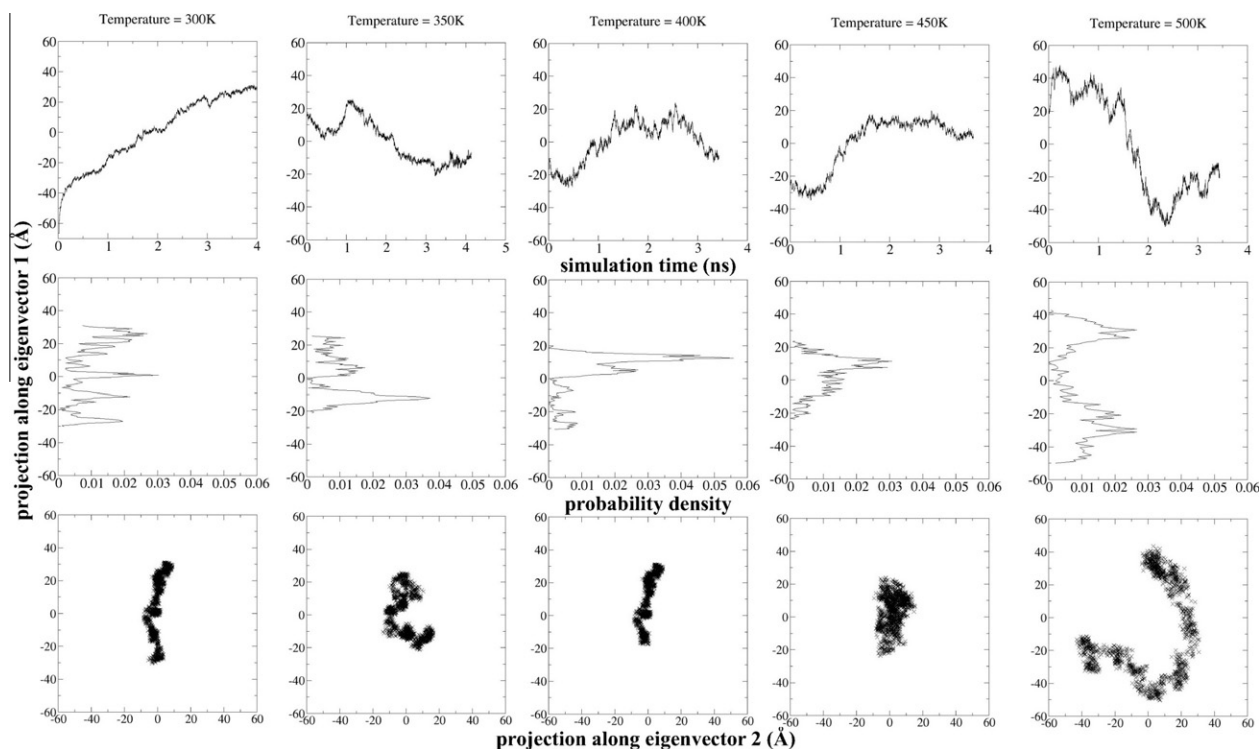


Fig. 12. Comparison of principal component analysis for 12 bp PNA strands in LDH (Model VI) as a function of temperature. The layout is identical to that used in Fig. 16. The first row displays the motion of heavy atoms along the first eigenvector, while the second row plots the probability density that the model occupies a particular configuration. The lowermost row shows the projections of the trajectory on planes defined by pairs of eigenvectors 1 and 2. Unlike the RNA in Models IV and VII, shown in Figs. 14 and 16, the PNA model has a more unimodal probability distribution up to 400 K and 50 atm, suggesting that thermal undulations of the LDH sheets do not affect the intercalated PNA as much as RNA and DNA, presumably because the peptide chains only interact weakly with the LDH sheets through van der Waals forces.

towards the LDH surface and hydrogen bonds between intra-strand base-pairs. The main contribution to the dynamics of the PNA strand differs from DNA/RNA, as the PNA has an uncharged peptide backbone which does not interact so strongly with the LDH surface. The peptide backbone oscillates on both sides towards/away from the

central axis. Animations of the first eigenvector of RNA, DNA and PNA intercalated and free in bulk water are provided in the [Electronic Annex](#). A superposition of configurations displaying the second eigenvector (second most dominant principal component) is provided in the [Electronic Annex](#).

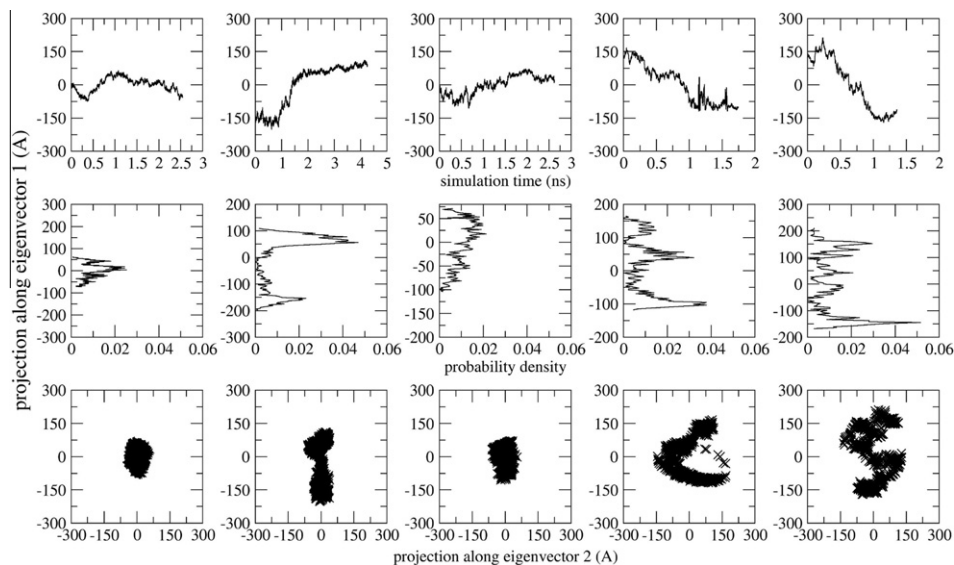


Fig. 13. Comparison of principal component analysis for 12 bp PNA strands in bulk water (Model IX) as a function of temperature. The layout is identical to that used in [Fig. 14](#). The first row displays the motion of heavy atoms along the first eigenvector, while the second row plots the probability density that the model occupies a particular configuration. The lowermost row shows the projections of the trajectory on planes defined by pairs of eigenvectors 1 and 2.

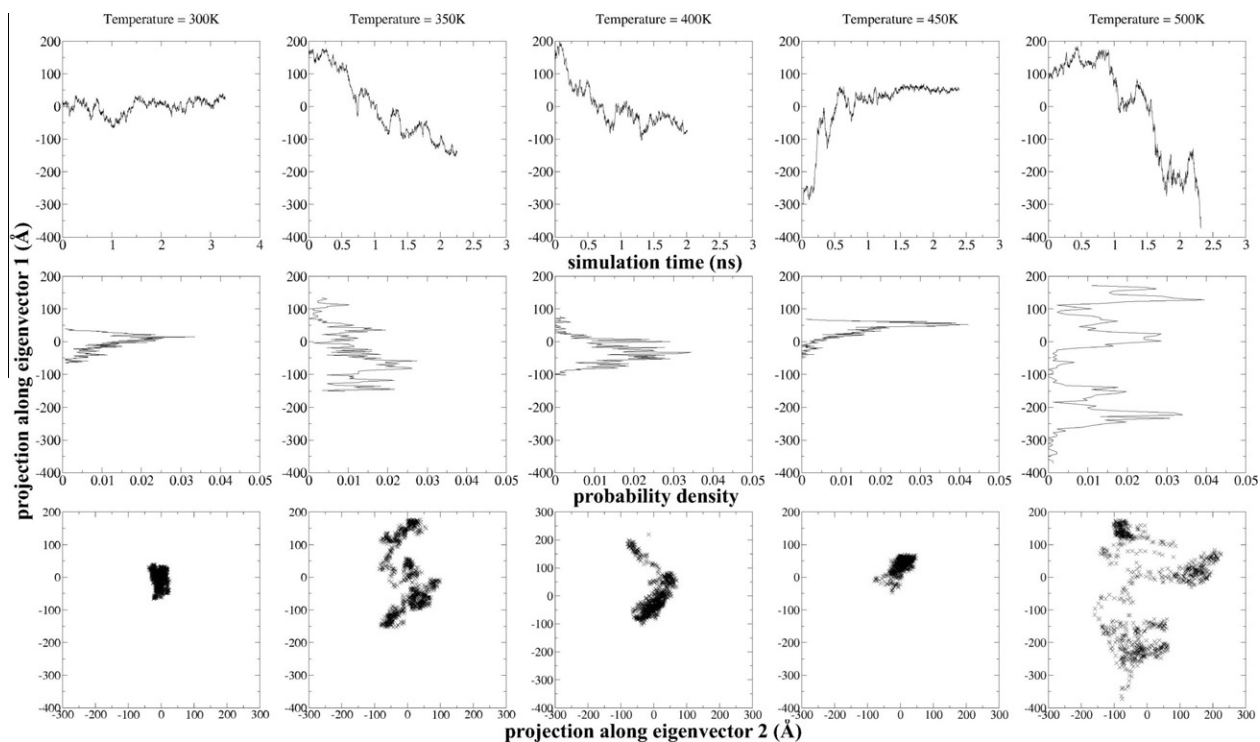


Fig. 14. Comparison of principal component analysis for 12 bp RNA strands in bulk water corresponding to Model VII at the temperatures indicated.

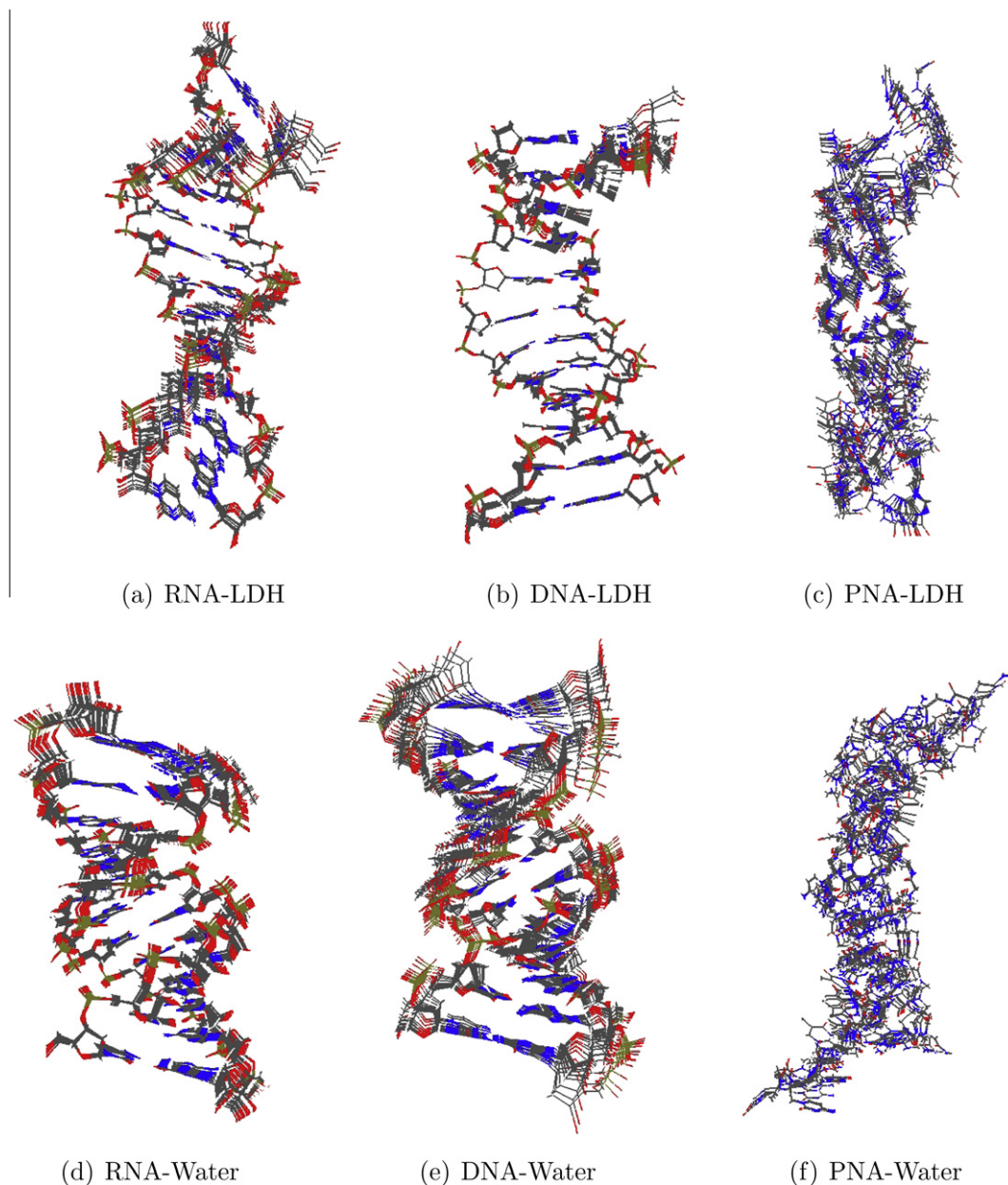


Fig. 15. Superposition of configurations obtained by projecting the motion of all atoms onto the first eigenvector using principal component analysis for (a) Model I; (b) Model II; (c) Model III; (d) Model VII; (e) Model VIII; and (f) Model IX. The configurations are averaged over the last nanosecond of simulation for the studied models. The colour scheme is the same as that used in Fig. 3. Intercalated nucleic acids show reduced motion compared to the corresponding bulk water models for the dominant mode of motion. DNA displays less motion, with respect to the first eigenvector, than all other models indicating that this is the most structurally stable of the models tested. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

As seen in previous studies (Suter et al., 2007; Thyveetil et al., 2007, 2008a,b), thermal undulations of LDH sheets are evident in large scale molecular dynamics simulations. These are found to significantly affect the motion of intercalated DNA strands (Thyveetil et al., 2008b,a) due to strong electrostatic interactions between the intercalant and the LDH sheet. PCA provides clear evidence that RNA strands are also greatly affected by the thermal motion of LDH sheets. Figs. 14 and 16 compares PCA results for intercalated RNA and RNA within bulk water. Under ambient

conditions, RNA intercalated within LDH as well as within bulk water has a unimodal probability distribution along the first eigenvector, suggesting that a well defined equilibrium structure has been reached. Figs. 14 and 16 show the subspace spanned by the first two principal components (the first principal component is shown in Fig. 15, the second is given in the [Electronic Annex](#)) for Models IV and VII at ambient temperatures and pressures. At higher temperatures and pressures, it becomes increasingly evident that projections along the first two eigenvector pairs are

coupled, which is most likely due to thermal motion in the LDH sheets increasingly influencing the motion of intercalated RNA (Thyveetil et al., 2008b). The main contribution to the dynamics at this hydration level is seen at the ends of the nucleic-acid molecules.

PNA intercalants are composed of an uncharged PNA protein backbone. In turn, this causes the dynamics of PNA to be different from duplex strands of RNA and DNA each of which possesses two charged complementary strands. At temperatures and pressures above ambient conditions, the number of retained Watson–Crick hydrogen bonds is larger than those found in intercalated 12 bp RNA, suggesting stronger base-pair interactions within the PNA strand. Basal spacing values for PNA–LDH are on the whole larger than RNA–LDH in Model IV, suggesting that the 12 bp RNA strand has less resilience to the LDH sheet dynamics and distorts more readily. The RMSD values for the PNA–LDH Model VI in Table 2 are larger than those for RNA–LDH in Model IV indicating that, overall, the PNA double strand deviates more from its starting structure at all temperatures and pressures compared to RNA, whilst retaining a higher proportion of Watson–Crick bonding. By contrast, the structural parameters of intercalated DNA with the same base-pair sequence, as detailed in Table 2, show that the DNA double helix retains its own integrity more than

the PNA duplex, relative to their individual starting structures. Table 2 shows that the LDH confers a different relative preferential stability on the three nucleic-acid duplexes compared to their homologues in bulk water. The relative trend in stability in terms of intact Watson–Crick base-pairs when intercalated is  $\text{PNA} < \text{RNA} < \text{DNA}$ . The reverse trend is observed for nucleic acids free in aqueous solution.

Although the structure of RNA is perturbed whilst intercalated, the RMSD and PCA data indicate that overall structure retention, at high temperatures and pressures, is similar to that exhibited at ambient conditions.

At ambient temperatures of 300 K PNA is found to have the smallest self-diffusion coefficient of the three intercalated nucleotides whilst RNA has the largest (shown in Table 3). The trend in diffusion coefficients is altered at elevated temperatures of 500 K where PNA has the highest and DNA the lowest self-diffusion coefficient.

#### 4. CONCLUSIONS

In this paper, we have undertaken a comparative study of the structure, properties and stability of selected double-stranded sequences of RNA, DNA and PNA. Our intention was to gain insight into which candidate genetic material arising at the time of the origin of life on Earth

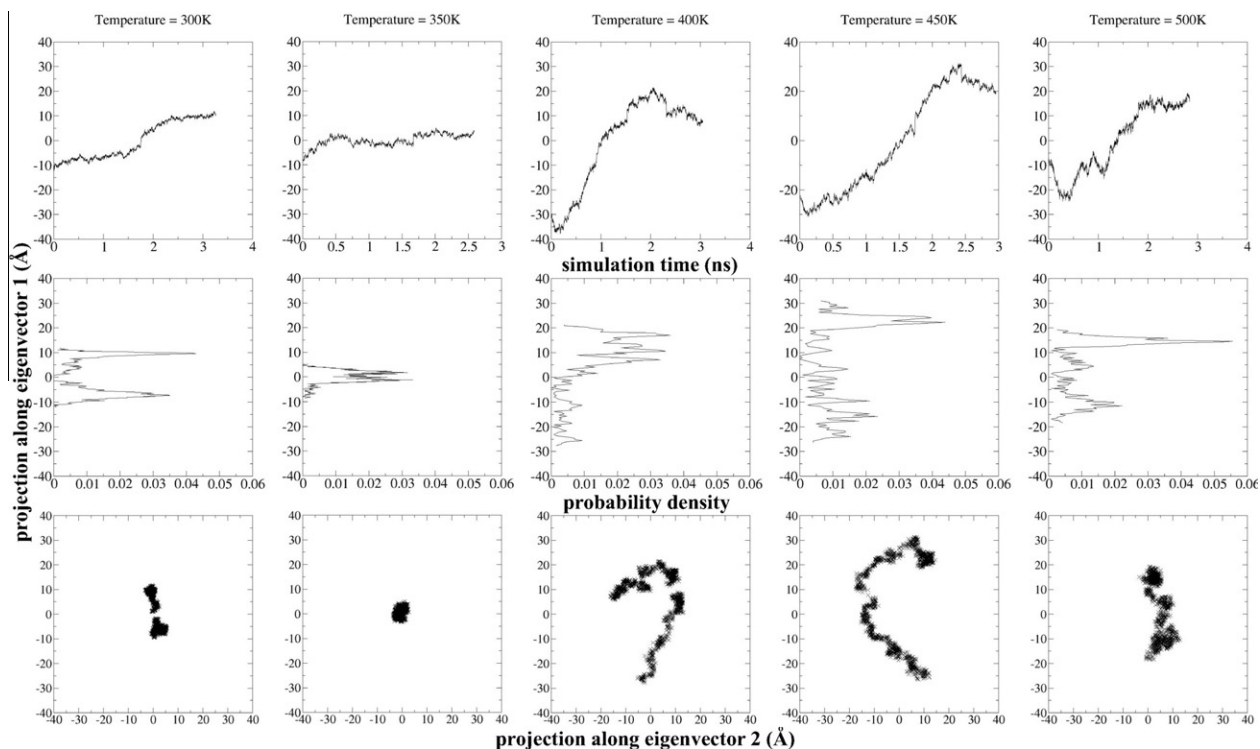


Fig. 16. Comparison of principal component analysis for 12 bp RNA strands in an LDH corresponding to Model IV at the temperatures indicated. The first row displays the motion of heavy atoms along the first eigenvector, while the second row exhibits the probability that the model occupies a particular configuration. The lowermost row shows the projections of the trajectory on planes defined by pairs of eigenvectors 1 and 2. Comparison of the axes scales shows that RNA intercalated within LDH has almost 10 times less degrees of freedom than RNA in bulk water (see Fig. 14). With increasing temperatures and pressures the strands have much larger amplitudes in their structural fluctuations; for example, consider the effect of heavy atom motion along the first eigenvector. The plots of the subspace occupied by the RNA atoms also appear more coupled when the RNA strands are intercalated in LDH compared to those in bulk water. This is evidence that the strands are influenced by the movement of the LDH sheets, as previously shown (Thyveetil et al., 2008b).



Table 3

Diffusion coefficients for 12 bp PNA oligomers in Model VI, RNA oligomers in Model IV and DNA oligomers in Model V calculated from the production phase of the simulations. Error bars are obtained from the least squares error fit on the gradient of the mean square displacement vs. time graph. As the temperature increases, the average diffusion within the basal plane increases.

Temperature (K)	PNA Diffusion coefficient ( $\text{\AA}^2 \text{ns}^{-1}$ )	RNA Diffusion coefficient ( $\text{\AA}^2 \text{ns}^{-1}$ )	DNA diffusion coefficient ( $\text{\AA}^2 \text{ns}^{-1}$ )
300	0.20 ± 0.000	0.33 ± 0.00	0.28 ± 0.01
350	0.90 ± 0.001	0.48 ± 0.00	0.55 ± 0.02
400	1.40 ± 0.002	0.83 ± 0.00	0.58 ± 0.03
450	2.60 ± 0.002	1.45 ± 0.00	1.80 ± 0.09
500	4.00 ± 0.004	2.34 ± 0.01	2.11 ± 0.12

may have been preferentially favoured by the prevailing geochemistry, in particular when interacting with anionic clays.

Study of the root-mean-squared deviation relative to the starting structures together with principal component analysis of the double-stranded nucleic acids indicate that the intercalated RNA, DNA and PNA molecules all have significantly reduced motion relative to the nucleic acids in bulk water. However, the relative similarity between starting and end-point structures does not offer the best insight into structure retention of each double stranded nucleotide. A more important measure of retention of structural integrity from a genetic information-transfer perspective is the number of Watson–Crick bonds maintained.

The simulations demonstrate that hydration plays an important role in determining the structural stability of all three intercalated nucleic acids. This is in agreement with our previous findings on DNA moieties within LDH (Thyveetil et al., 2008b). The results of our present study show that the three nucleic acids are affected differently by the degree of hydration of the LDH. While RNA continues to adopt a structure closer to the crystalline A-form at all levels of hydration, at lower hydration levels there is also a tendency for DNA to adopt structures closer to the A-form than the B-form found in bulk water. With regard to Watson–Crick hydrogen-bonding, the DNA duplexes retain a greater structural integrity as compared to the intercalated RNA and PNA double strands which manifest significant degradation in base-pairing.

Having examined the properties of selected 12-mer duplexes of RNA, DNA and PNA in bulk water and when intercalated in the interlayer of LDH minerals, together with the hydration properties of the nucleic acid–LDH complexes, we arrive at the conclusion that the properties of the DNA, RNA and PNA duplexes are strongly dependent on the state of hydration of the LDH interlayers. For the assembly and maintenance of the first double strand genetic system, a hydrated environment is necessary and, therefore, it is vital for there to be water within the LDH interlayers.

All of the nucleic acids retain some degree of structural integrity when intercalated in LDH minerals, even at elevated temperatures and pressures. This supports the notion that such anionic clays, possibly present within early Archean hydrothermal systems as Fe(II)/Fe(III) green rusts, may have played a significant role in concentrating and catalysing early prebiotic chemistry, as proposed by Arrhenius (2003).

LDH-intercalated 12 bp double-stranded RNA oligomers were generally found to deform from their initial structure more readily compared to DNA and PNA. In addition, the 12 bp RNA and PNA strands are more easily deformed from their initial starting structure in bulk water compared to DNA; significantly more water is needed within LDH sheets to retain the double helix shape of *ds*-RNA and *ds*-PNA compared to the situation in bulk water, while there is a noticeable preference for intercalated DNA to retain an A-form, especially at lower levels of hydration.

To get a proper handle on the behaviour of these nucleotides within clays, a range of analysis tools needs to be brought to bear on the molecular dynamics trajectory data. Though the nucleic acid strands may lose their Watson–Crick base pairing in layered double hydroxides, the root-mean-squared deviation relative to the initial nucleic acid structure is invariably smaller for the intercalated nucleic acid than for the nucleic acids in bulk water. The intercalated species, stabilized through strong Coulombic interactions with the layered double hydroxides sheets, is unable to distort much from its initial coordinates; instead the Watson–Crick base-pairing becomes disrupted at increased temperature and pressure, assisted by the thermal motions of the layered double hydroxides sheets.

Of the three nucleic acids under investigation, DNA is observed to be the most stable when intercalated within an LDH, with respect to the percentage of Watson–Crick base-pairs. The opposite is found in bulk water where both PNA and RNA are more stable than DNA. The uncharged protein backbone of PNA has a detrimental effect on the overall stability of the polymer when intercalated as a duplex, as it experiences a greatly reduced electrostatic interaction with the charged layered double hydroxides sheets. These results indicate that a mineral based origin of life may well have been rather different from the aqueous, bulk water based one more commonly considered in origins of life scenarios, DNA being the most stable genetic material within our simulations. Our findings further suggest that a mineral mediated origin of life may have favoured DNA as the informational storage biomolecule over competing RNA and PNA, providing a route to modern biology from the RNA world.

#### ACKNOWLEDGMENTS

This research was supported by the National Science Foundation through TeraGrid/XSEDE resources provided by the Pittsburgh Supercomputing Center (Bigben), the Texas Advanced

Computing Center (Ranger) and the National Institute of Computer Science (Kraken) within allocations under NRAC Grant MCA04N014 and MRAC Grant DMR070013N, and by resources of the Argonne Leadership Computing Facility (Intrepid) at Argonne National Laboratory, supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357. The authors are also indebted to the UK National Grid Service for access to 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 machine. The authors also thank Sarah Harris for in-depth discussions on nucleic acid simulation. J.B.S. is funded by an EPSRC Ph.D. studentship.

## APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gca.2011.12.023.

## REFERENCES

- Amadei A., Linssen A. B. M. and Berendsen H. J. C. (1993) Essential dynamics of proteins. *Proteins* **17**, 412–425.
- Arrhenius G. (1993) *Phosphate in Models for Chemical Evolution*. Deepak Publishers, Hampton, Virginia, USA.
- Arrhenius G. (2003) Crystals and life. *Helv. Chim. Acta* **86**, 1569–1586.
- Arrhenius G., Sales B., Mojszsis S. and Lee T. (1997) Entropy and charge in molecular evolution – the case of phosphate. *J. Theor. Biol.* **187**(4), 503–522.
- Bean H. D., Anet F. A. L., Gould I. R. and Hud N. V. (2006) Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA. *Origins Life Evol. Biosph.* **36**(1), 39–63.
- Bellotto M., Rebours B., Clause O., Lynch J., Bazin D. and Elkaim E. (1996) A reexamination of hydrotalcite crystal chemistry. *J. Phys. Chem.* **100**(20), 8527–8534.
- Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F. and Hermans J. (1981) *Intermolecular Forces*, first ed. Reidel Publishing Co..
- Boek E. S., Coveney P. V. and Skipper N. T. (1995a) Molecular modeling of clay hydration: a study of hysteresis loops in the swelling curves of sodium montmorillonites. *Langmuir* **11**(12), 4629–4631.
- Boek E. S., Coveney P. V. and Skipper N. T. (1995b) Monte Carlo molecular modeling studies of hydrated Li-, Na-, and K-smectites: understanding the role of potassium as a clay swelling inhibitor. *J. Am. Chem. Soc.* **117**(50), 12608–12617.
- Budin I. and Szostak J. (2010) Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* **39**, 245–263.
- Case D. A., Darden T. A., Cheatham, III, T. E., Simmerling C. S., Wang J., Duke R. E., Luo R., Walker R. C., Zhang W., Merz K. M., Roberts B. P., Wang B., Hayik S., Roitberg A., Seabra G., Kolossvai I., Wong K. F., Paesani F., Vanicek J., Liu J., Wu X., Brozell S. R., Steinbrecher T., Gohlke H., Cai Q., Ye X., Wang J., Hsieh M.-J., Cui G., Roe D. R., Mathews D. H., Seetin M. G., Sagui C., Babin V., Luchko T., Gusarov S., Kovalenko A. and Kollman P. A. (2011) *AMBER 11*. University of California, San Francisco.
- Cheatham, III, T. E. and Young M. A. (2001) Molecular dynamics simulation of nucleic acids: successes, limitations, promise. *Biopolymers* **56**, 232–256.
- Cheng L. and Unrau P. (2010) Closing the circle: replicating RNA with RNA. *Cold Spring Harb. Perspect. Biol.* **2**, a002204.
- Choy J. H., Choi S. J., Oh J.-M. and Park T. (2007) Clay minerals and double layered hydroxides for novel biological applications. *Appl. Clay Sci.* **36**, 122–132.
- Choy J.-H., Kwak S.-Y., Jeong Y.-J. and Park J.-S. (2000) Inorganic layered double hydroxides as nonviral vectors. *Angew. Chem.* **39**(22), 4041–4045.
- Choy J.-H., Kwak S.-Y., Park J.-S. and Jeong Y.-J. (2001) Cellular uptake behavior of [ $\gamma$ - $^{32}$ P] labeled ATP–LDH nanohybrids. *J. Mater. Chem.* **11**, 1671–1674.
- Choy J.-H., Kwak S.-Y., Park J.-S., Jeong Y.-J. and Portier J. (1999) Intercalative nanohybrids of nucleoside monophosphates and DNA in layered metal hydroxide. *J. Am. Chem. Soc.* **121**(6), 1399–1400.
- Choy J.-H., Oh J.-M., Park M., Sohn K.-M. and Kim J.-W. (2004) Inorganic–biomolecular hybrid nanomaterials as a genetic molecular code system. *Adv. Mater.* **16**(14), 1181–1184.
- Choy J.-H. and Son Y.-H. (2004) Intercalation of vitamer into Idh and their controlled release properties. *Bull. Korean Soc.* **25**(1), 122–126.
- Cormack A. N., Lewis R. J. and Goldstein A. H. (2004) Computer simulation of protein adsorption to a material surface in aqueous solution: biomaterials modeling of a ternary system. *J. Phys. Chem. B* **108**, 20408–20418.
- Cygan R. T., Liang J.-J. and Kalinichev A. G. (2004) Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *J. Phys. Chem. B* **108**, 1255–1266.
- de Groot B. L., Daura X., Mark A. E. and Grubmüller H. (2001) Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. *J. Mol. Biol.* **309**, 299–313.
- Desigaux L., Belkacem M. B., Richard P., Cellier J., Leone P., Cario L., Leroux F., Taviot-Gueho C. and Pitard B. (2006) Self-assembly and characterization of layered double hydroxide/DNA hybrids. *Nano Lett.* **6**(2), 199–204.
- Duffy D. M. and Harding J. H. (2004) Simulation of organic monolayers as templates for the nucleation of calcite crystals. *Langmuir* **20**, 7630–7636.
- Follmann H. and Brownson C. (2009) Darwin’s warm little pond revisited: from molecules to the origin of life. *Naturwissenschaften* **96**, 1265–1292.
- Fu Q., Sherwood Lollar B., Horita J., Lacrampe-Couloume G. and Seyfried, Jr., W. E. (2007) Abiotic formation of hydrocarbons under hydrothermal conditions: constraints from chemical and isotope data. *Geochim. Cosmochim. Acta* **71**, 1982–1998.
- Greenwell H. and Coveney P. V. (2006) Layered double hydroxide minerals as possible prebiotic information storage and transfer compounds. *Origins Life Evol. Biosph.* **36**, 13–37.
- Greenwell H., Jones W., Coveney P. V. and Stackhouse S. (2006) On the molecular modelling of the structure and properties of clays: a materials chemistry perspective. *J. Mater. Chem.* **16**, 708–723.
- Hansma H. G. (2010) Possible origin of life between mica sheets. *J. Theor. Biol.* **266**, 175–188.
- Harding J. H. and Duffy D. M. (2006) The challenge of biominerals to simulations. *J. Mater. Chem.* **16**, 1105–1112.
- Hazen R. and Sverjensky D. (2010) Mineral surfaces, geochemical complexities, and the origins of life. *Cold Spring Harb. Perspect. Biol.* **2**, a002162.
- Hess B. (2000) Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E* **62**, 8438–8448.
- Hobza P., Kabelac M., Sponer J. and Mejzlík J. V. (1997) Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum

- chemical methods (AM1, MNDO/M, PM3), and ab initio Hartree–Fock method for interaction of DNA bases: comparison with nonempirical beyond Hartree–Fock results. *J. Comp. Chem.* **18**, 1136–1150.
- Jeffares D., Poole A. and Penny D. (1998) Relics from the RNA world. *J. Mol. Evol.* **46**(1), 18–36.
- Jha S., Coveney P. V. and Laughton C. A. (2005) Force field validation for nucleic acid simulations: comparing energies and dynamics of a DNA dodecamer. *J. Comp. Chem.* **26**, 1617–1627.
- Joshi P., Aldersley M., Delano J. and Ferris J. (2009) Mechanism of montmorillonite catalysis in the formation of RNA oligomers. *J. Am. Chem. Soc.* **131**(37), 13369–13374.
- Joyce G. (2002) Booting up life. *Nature* **418**, 214–221.
- Krishnamurthy R., Pitsch S. and Arrhenius G. (1996) Mineral induced formation of pentose-2,4-bisphosphates. *Origins Life Evol. Biosph.* **26**(3), 240–241.
- Kwak S.-Y., Kriven W. M., Wallig M. A. and Choy J.-H. (2004) Inorganic delivery vector for intravenous injection. *Biomaterials* **25**(28), 5995–6001.
- Kwak S.-Y., Park J.-S., Jeong Y.-J. and Choy J.-H. (2002) Bio-LDH nanohybrid for gene therapy. *Solid State Ionics* **151**(1–4), 229–234.
- Lu X.-J. and Olson W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121.
- Lu X.-J. and Olson W. K. (2007) A software package for the analysis, rebuilding, and visualization of three-dimensional nucleic acid structures. Available from: <<http://rutchem.rutgers.edu/~xiangjun/3DNA/>>.
- Martin W., Baross J., Kelley D. and Russell M. (2008) Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805–814.
- Martin W. and Russell M. J. (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos. Trans. Biol. Sci.* **358**, 59–85.
- Mathew D. and Luthey-Schulten Z. (2010) Influence of montmorillonite on nucleotide oligomerization reactions: a molecular dynamics study. *Orig. Life Evol. Biosph.* **40**, 303–317.
- McCollom T. M. and Shock E. L. (1997) Hydrothermal geochemistry and the deep biosphere. *Geochim. Cosmochim. Acta* **61**, 4375–4391.
- Nelson K., Levy M. and Miller S. (2000) Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. USA* **97**(8), 3868.
- Nielsen P. and Egholm M., et al. (1999) An introduction to peptide nucleic acid. *Curr. Issues Mol. Biol.* **1**(1–2), 89–104.
- Nielsen P. E. (2007) Peptide nucleic acids (PNA) in chemical biology and drug discovery. *Chem. Biodivers.* **4**, 1996–2002.
- Nielsen P. E., Egholm M., Berg R. H. and Buchardt O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254**, 1497.
- Oh J.-M., Kwak S.-Y. and Choy J.-H. (2006) Intracrystalline structure of DNA molecules stabilized in the layered double hydroxide. *J. Phys. Chem. Solids* **67**(5–6), 1028–1031.
- Orgel L. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123.
- Pitsch S., Eschenmoser A., Gedin B., Hui S. and Arrhenius G. (1995) Mineral induced formation of sugar phosphates. *Orig. Life Evol. Biosph.* **25**(4), 297–334.
- Plimpton S. (1995) Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.* **117**, 1–19.
- Poole A., Jeffares D. and Penny D. (1998) The path from the RNA world. *J. Mol. Evol.* **46**(1), 1–17.
- Powner M., Gerland B. and Sutherland J. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242.
- Scott W. G., Murray J. B., Arnold J. R. P., Stoddard B. L. and Klug A. (1996) Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science* **274**, 2065–2069.
- Shields G., Laughton C. and Orozco M. (1998) Molecular dynamics simulation of a PNA DNA PNA triple helix in aqueous solution. *J. Am. Chem. Soc.* **120**, 5895–5904.
- Soliva R., Sherer E., Luque F., Laughton C. and Orozco M. (2000) Molecular dynamics simulations of PNA–DNA and PNA–RNA duplexes in aqueous solution. *J. Am. Chem. Soc.* **122**, 5997–6008.
- Suter J. L., Coveney P. V., Greenwell H. C. and Thyveetil M.-A. (2007) Large-scale molecular dynamics study of montmorillonite clay: emergence of undulatory fluctuations and determination of material properties. *J. Phys. Chem. C* **111**, 8248–8259.
- Sutherland J. (2010) Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* **2**, a005439.
- Swadling J. B., Coveney P. V. and Greenwell H. C. (2010) Clay minerals mediate folding and regioselective interactions of RNA: a large-scale atomistic simulation study. *J. Am. Chem. Soc.* **132**, 13750–13764.
- Thyveetil M., Coveney P. V., Greenwell H. and Suter J. (2008a) Role of host layer flexibility in DNA guest intercalation revealed by computer simulation of layered nanomaterials. *J. Am. Chem. Soc.* **130**, 12485–12495.
- Thyveetil M.-A., Coveney P. V., Greenwell H. and Suter J. (2008b) Computer simulation study of the structural stability and materials properties of DNA-intercalated layered double hydroxides. *J. Am. Chem. Soc.* **130**, 4742–4756.
- Thyveetil M.-A., Coveney P. V., Suter J. L. and Greenwell H. C. (2007) Emergence of undulations and determination of materials properties in large-scale molecular dynamics simulation of layered double hydroxides. *Chem. Mater.* **19**, 5510–5523.
- Unrau P. J. and Bartel D. P. (1998) Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *Nature* **395**, 260–263.
- Ura Y., Beierle J., Leman L., Orgel L. and Ghadiri M. (2009) Self-assembling sequence-adaptive peptide nucleic acids. *Science* **325**, 73.
- Wang J., Cieplak P. and Kollman P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?. *J. Comp. Chem.* **21**, 1049–1074.
- Wang J., Kalinichev A. G., Kirkpatrick R. J. and Hou X. (2001) Molecular modeling of the structure and energetics of hydroxalite hydration. *Chem. Mater.* **13**, 145–150.
- Woese C. (1967) *The Genetic Code: The Molecular Basis for Denetic Expression*, first ed. Harper & Row, New York.
- Wu M. and Higgs P. (2009) Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the RNA world. *J. Mol. Evol.* **69**, 541–554.
- Young M. A., Jayaram B. and Beveridge D. L. (1997) Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* **119**, 59–69.