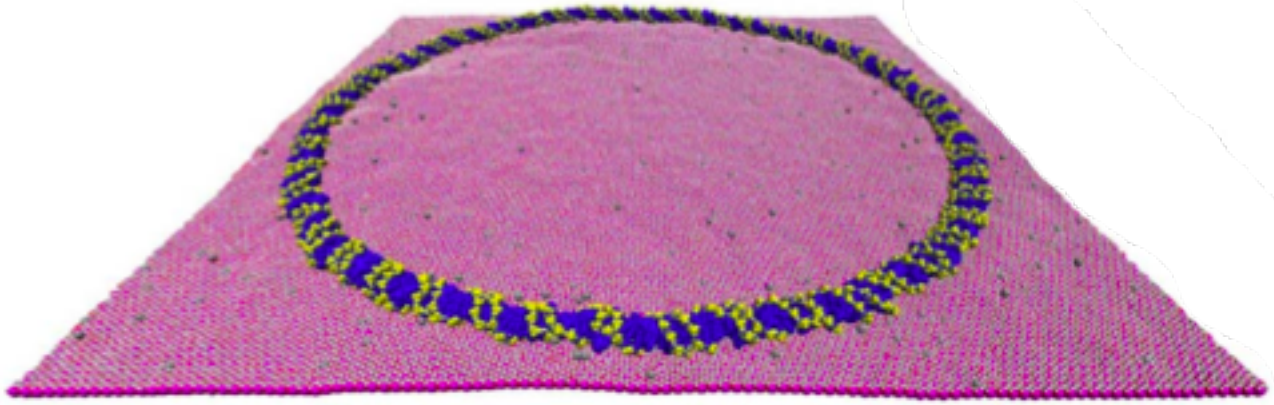
UCL

Theory, Modelling and Simulation in Origins of Life Studies

Peter V. Coveney, Jacob B Swadling, Jonathan AD Wattis and H Christopher Greenwell



Contents

Current theories in origins of life studies

The RNA world

The RNA world scenario claims that RNA was the key biopolymer in the first life because RNA can serve both as a catalyst as well as repository of genetic information.

Indirect evidence for the RNA world:

- Cech & Altman won the 1989 Nobel prize for the independent discovery of the group of catalytic RNA molecules known as ribozymes, which were subsequently shown to catalyse a significant number of reactions.

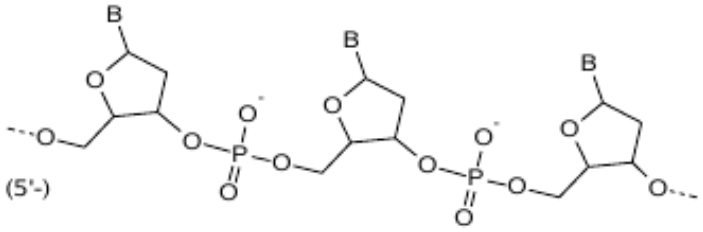
- Steitz, who won the 2009 Nobel prize in chemistry, along with Moore who determined the structure of the ribosome, showing that it is a ribozyme.

T. R. Cech (1986) *PNAS*, 83:4360-4363.
T. A. Steitz *et al.* (2000) *Science*, 289:905.

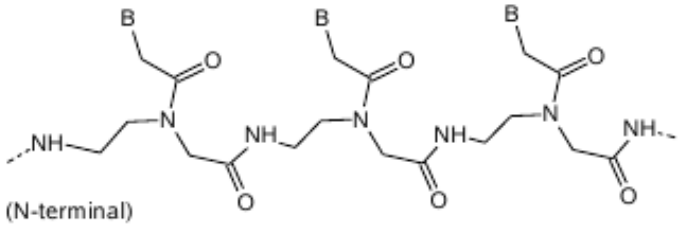
Current theories in origins of life studies

Unanswered questions:

- How did self replicating molecules get going?
- How, and at what point, did chiral selection occur?
- Was there a pre-RNA world?



DNA



PNA

Peptide nucleic acid (PNA) is an artificial nucleic acid. PNA's backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds. The various purine and pyrimidine bases are linked to the backbone by methylene carbonyl bonds.

It has been hypothesized that the earliest life on Earth may have used PNA as a genetic material due to its extreme robustness, simpler formation and possible spontaneous polymerization at 100°C

P. E. Nielson *et al.* (1991) *Science,* 254:1497-500.

Current theories in origins of life studies

Primordial soup

First proposed by Haldane in 1929. UV radiation provided energy to convert methane, ammonia and water into the first organic molecules. These molecules were concentrated in a "hot organic soup".
Miller and Urey (1959) produced amino acids by releasing an electrical discharge through a mixture of water/methane/ammonia/hydrogen.
Recent evidence supporting the primordial soup came from Sutherland who synthesised activated pyrimidine nucleotides in the "soup" conditions.

Problems:

- Absence of geochemical evidence.
- Ionizing UV radiation inherently destroys as much as is creates.
- The "soup" is usually conceived to be in thermodynamic equilibrium; there is no free energy to allow further reaction.

Haldane, J. B. S. *et al.* (1929) *Rationalist Annual* 3: 3-10
Sutherland J. D. *et al.* (2009) *Nature* 459: 239-42
Stanley, M. L. and Urey, H. C. (1959). Science 130: 245.

Current theories in origins of life studies

Alkaline Vents

A type of alkaline vent was discovered in 2000 by Kelley. These alkaline vents form from reaction of seawater with minerals like olivine as a result of serpentinisation. Moderately high temperatures (~473K), alkaline (pH 9-11), and rich in hydrogen.

Lost City Alkaline hydrothermal vent found off the Mid-Atlantic Ridge

- Olivine is a magnesium iron silicate, and is one of the most abundant minerals on Earth.
- Serpentinisation is the geochemical process in which olivine is hydroxylated to serpentine. The hydroxylated rocks expand and fracture, allowing entry of more seawater, perpetuating the reaction.

Kelley, D. S. *et al.* (2001) *Nature.* 412: 145-9
Russell, M. J. *et al.* (1994) *J. Mol. Evol.* 39: 231-43

Hypercycles

Some of the most original theoretical modelling of the origin and evolution of life came from Eigen and Schuster in 1977; in their three part paper titled “The hypercycle: A principal of natural self-organization”.

Eigen, M. and Schuster, P. (1977) *Naturwissenschaften* 11: 541-565

Self replicative macromolecules, such as RNA and DNA in a suitable environment, exhibit a behaviour which we may call “Darwinian” and which can be formally represented by the concept of the “quasi species”*.

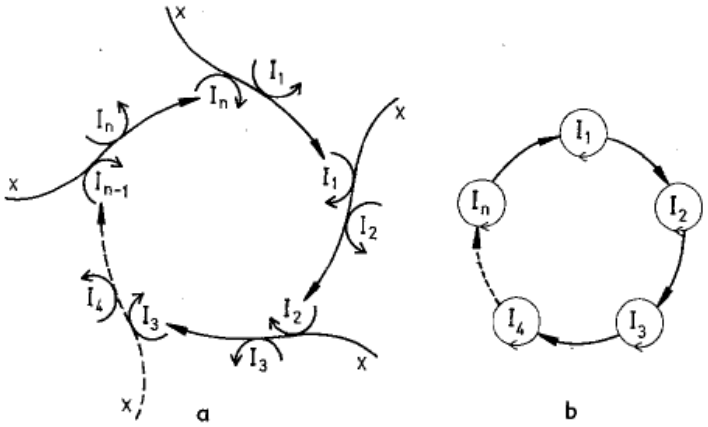


Fig. 7. *A catalytic hypercycle* consists of self-instructive units I_i with two-fold catalytic functions. As autocatalysts or—more generally—as catalytic cycles the intermediates I_i are able to instruct their own reproduction and, in addition, provide catalytic support for the reproduction of the subsequent intermediate (using the energy-rich building material X). The simplified graph (b) indicates the cyclic hierarchy

© P.V. Coveney *et al.*

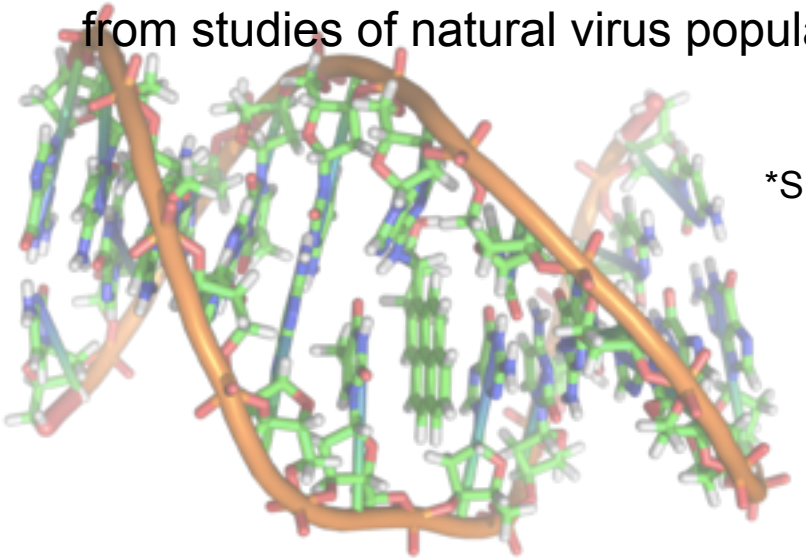
* Hypercycles shown to be unstable without spatial compartmentalisation (i.e. “cells”).
Maynard-Smith, J. (1979) *Nature* 280: 445-446

Hypercycles

- The quasi-species itself represents the stationary distribution of macromolecular sequences maintained by chemical reactions effecting error-prone replication, and by transport processes.

- The quasi-species model demonstrates how macromolecular information originates through non-equilibrium auto-catalytic reactions and thus forms a bridge between reaction kinetics and molecular evolution.

- Experimental data* obtained from test-tube evolution of polynucleotides and from studies of natural virus populations support the quasi-species model.



*Spiegelman, S. Q. (1971) *Rev. Biol.* 4: 213

Hypercycles

- Non-equilibrium processes are of central importance in both producing life and sustaining it.
- Life is not just about replication; it is also a coupling of chemical reactions.
- Life needs a continuous and replenishing source of chemical energy.

© P.V. Coveney *et al.*

Modelling elements of the RNA-world

Analyzing a microscopic kinetic model for the emergence of long chains of RNA from monomeric β-D-ribonucleotide precursors in prebiotic circumstances.

The models incorporate the possibility of:

(i) direct chain growth

$$C_r^{\gamma} + N_i \overset{slow}{\rightleftharpoons} C_{r+1}^{\gamma+N_i}.$$

(ii) template assisted synthesis

$$C_r^{\gamma} + N_i + C_s^{\theta} \overset{fast}{\rightleftharpoons} C_{r+1}^{\gamma+N_i} + C_s^{\theta},$$

(iii) ribozymal catalysis

$$C_r^{\gamma} + N_i + C_{r+k}^{\gamma*+\theta*} + C_s^{\xi} \rightleftharpoons C_{r+1}^{\gamma+N_i} + C_{r+k}^{\gamma*+\theta*} + C_s^{\xi}.$$

J. A. D. Wattis and P. V. Coveney. The origin of the RNA world: a kinetic model. *Journal of Physical Chemistry B*, 103:4231–4250, 1999.

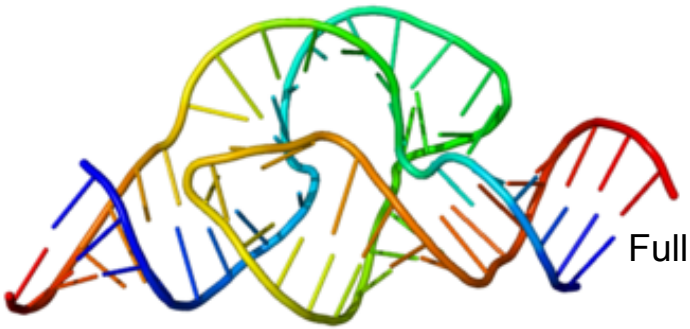
Modelling elements of the RNA-world

(iv) hydrolysis of polymerizing RNA sequences:

$$C_{r+s}^{\gamma+\theta} \rightarrow C_r^{\gamma} + C_s^{\theta}.$$

Conclusion

It is possible to realize the selection of certain self-replicating RNA polymer chains in a reasonable amount of time starting from plausible assumptions about the chemistry and initial conditions that could have prevailed within a putative prebiotic soup comprised of β-D-ribonucleotide monomers.



Full length hammerhead ribozyme.

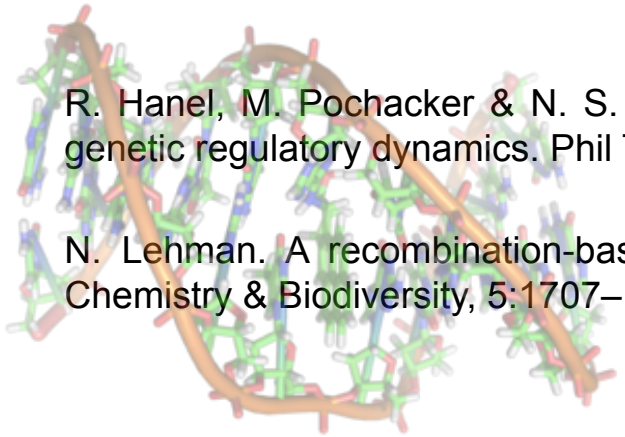
J. A. D. Wattis and P. V. Coveney. The origin of the RNA world: a kinetic model Journal of Physical Chemistry B, 103:4231–4250, 1999.

Modelling elements of the RNA-world

With small numbers of molecules involved, more inherently probabilistic models have an important role to play:

- Hanel *et al.* consider the kinetics of stochastic minimally nonlinear models.

- Lehman proposes that self-replication occurred via RNA strands breaking into shorter strands which replicated the daughter strands, recombining to form multiple copies of the original RNA molecule.

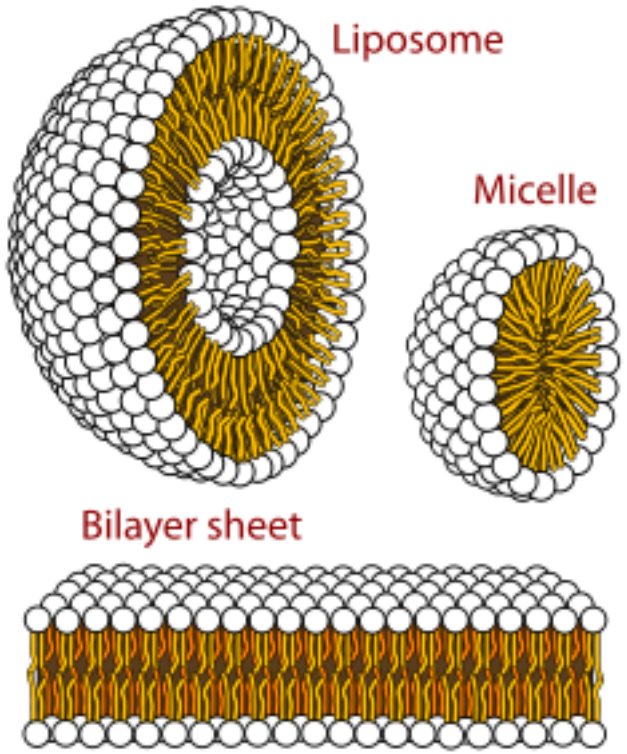


R. Hanel, M. Pochacker & N. S. Thurner. Living on the edge of chaos: minimally nonlinear models of genetic regulatory dynamics. Phil Trans Roy Soc A, 368: 5583–5596, 2010.

N. Lehman. A recombination-based model for the origin and early evolution of genetic information. Chemistry & Biodiversity, 5:1707–1717, 2008

Compartmentalisation

- Experimental scenarios leading to self-reproducing vesicles have been successfully developed by Luisi's group.
- Coveney and Wattis constructed a kinetic model which describes the stepwise growth and fragmentation of vesicular structures.
- Szostack has performed experiments on the replication of RNA inside vesicles.



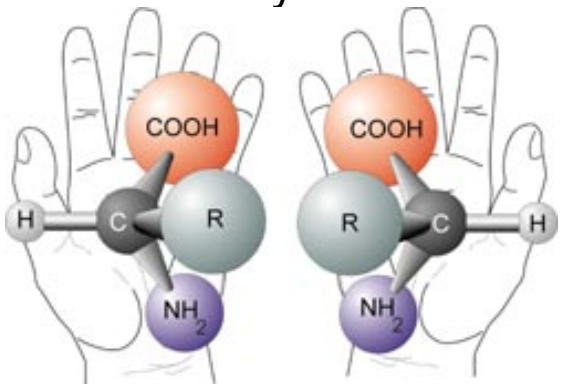
P. Walde, R. Wick, M. Fresta, A. Mangone, P. L. Luisi. J Am Chem Soc, 116: 11649–11654, 1994.

P. V. Coveney and J. A. D. Wattis. J. Chem. Soc.: Faraday Transactions, 102: 233–246, 1998.

A. I. Chen, K. Salehi-Ashtiani, J. W. Szostak. J Am Chem Soc. 127: 13213-9, 2005.

Homochiral crystals

- The first experiments illustrating symmetry-breaking in crystallisation were those of Kondepudi *et al.*
- With no stirring, approximately equal amounts of left- and right-handed crystals are formed.
- Noorduin *et al.* showed that grinding a mixture of chiral crystals eventually leads to a homochiral distribution of crystals of amino acids.



D. K. Kondepudi, R. J. Kaufman & N. Singh Science, 250: 975–976, 1990.
D. K. Kondepudi, K. L. Bullock, J. A. Digits P. D. Yarborough. J Am Chem Soc, 117:401–404, 1995.
W. L. Noorduin, T. Izumi, A. Millemaggi, M. Leeman, H. Meekes, W. J. P. van Enckevort, R. M. Kellogg, B. Kaptein, E. Vlieg & D. G. Blackmon. J Am Chem Soc, 130:1158–1159, 2008.

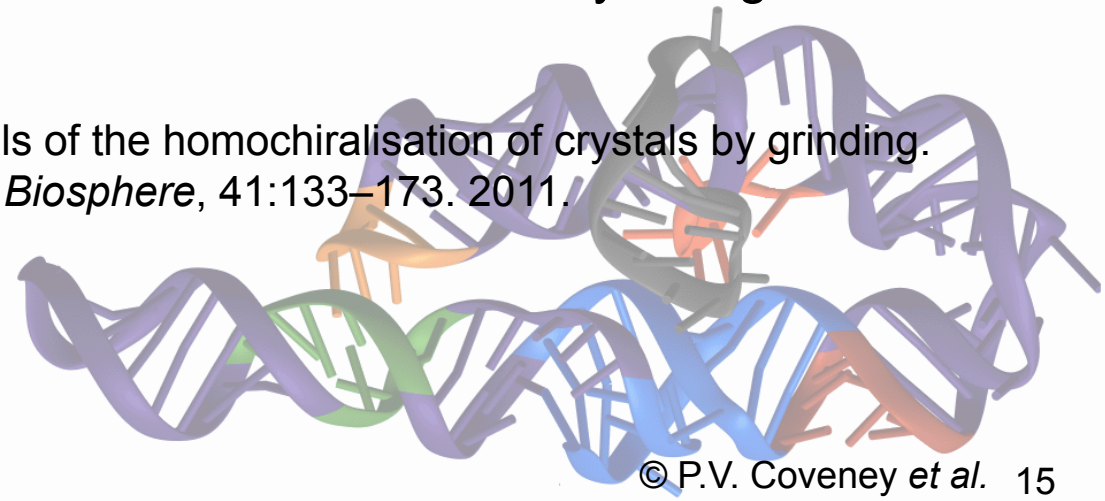
Modeling homochiralisation

- Before any experimental evidence was published, Uwaha had written down a crude model describing the processes of homochiralisation through grinding.

M. Uwaha. A model for complete chiral crystallization. J Phys Soc Jap, 73:2601–2603, 2004.

- Wattis describes a sequence of models starting from a detailed microscopic model which includes standard crystal growth and dissolution.

J. A. D. Wattis. Mathematical models of the homochiralisation of crystals by grinding. *Origins of Life and Evolution of the Biosphere*, 41:133–173. 2011.



© P.V. Coveney *et al.*

Homochiral polymerisation

Sanders' Model

This toy model has many features in common with Frank's 1953 paper on spontaneous asymmetric synthesis, but is adapted to an early peptide world.

P. G. H. Sandars. A toy model for the generation of homochirality during polymerization. Origins of Life and Evolution of Biospheres, 33: 575–587, 2003.

$$
\begin{aligned}
S \to L_1, \qquad & S \to R_1, \qquad && \text{slow}, \\
S+L_n \to L_1+L_n, \qquad & S+R_n \to R_1+R_n, \qquad && \text{autocatalytic, rate} \propto 1+f, \\
S+R_n \to L_1+R_n, \qquad & S+L_n \to R_1+L_n, \qquad && \text{cross-catalytic, rate} \propto 1-f,
\end{aligned}
$$

Long left-handed chains L_n catalyse the production of left-handed monomers L_1 from S, as well as the production of right-handed monomers, R_1.

JAD Wattis & PV Coveney. *Orig. Life Evol. Bios*. 35: 243-273, (2005)

Homochiral polymerisation

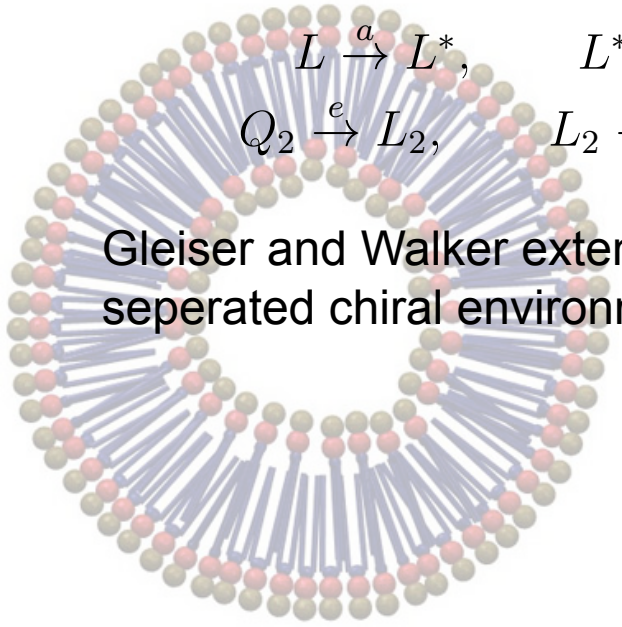
Plasson's scheme

Their scheme has monomers of each chirality (*L, R*) which can be *activated* through an input of energy into the system (to *L*, R**).

R. Plasson, H. Bersini & A. Commeyras. Recycling Frank: spontaneous emergence of homochirality in noncatalytic systems. Proc Natl Acad Sci, 101:16733–16738, 2004.

The processes in Plasson et. al.'s scheme can be summarised by

$$L \xrightarrow{a} L^*, \qquad L^* \xrightarrow{b} L, \qquad L^* + L \xrightarrow{p} L_2, \qquad L^* + R \xrightarrow{\alpha p} Q_2,$$
$$Q_2 \xrightarrow{e} L_2, \qquad L_2 \xrightarrow{\gamma e} Q_2, \qquad L_2 \xrightarrow{h} L + L, \qquad Q_2 \xrightarrow{\beta h} L + R,$$



Gleiser and Walker extended Plasson's scheme to describe protocells (spatially seperated chiral environments).

M. Gleiser and S. I. Walker. *Orig. Life Evol. Bios*. 39: 479-493, (2009).

Molecular simulation methods

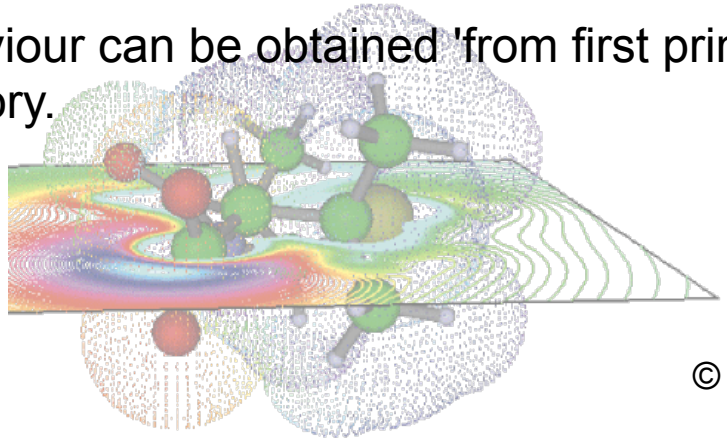
To model the structure and interaction of clay/organic systems we also employ other simulation methods including:

Molecular Dynamics: MD is a computer simulation technique where the time evolution of a set of atoms is followed by integrating their equations of motion.

Monte Carlo: computational algorithms that rely on repeated random sampling to compute their results.

Quantum Chemistry: Used to calculate the charge distribution of (small) molecules.

Ab Initio MD: Electronic behaviour can be obtained 'from first principles' by using Density Functional Theory.



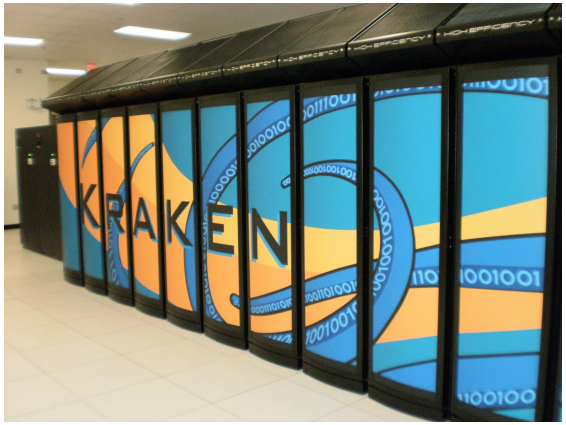
Molecular simulation methods

Notable resources used in our current research:

Ranger, at Texas Advanced Computing Centre (TACC), on the TeraGrid.
Sun Constellation machine with 62,976 compute cores, 123 TB of total memory and 1.7 PB of raw global disk space. 0.43 Petaflops (#17 Top500)

Kraken, at National Institute for Computational Science (NICS), on the TeraGrid.
XT5 machine with 99,072 cores. A peak performance of 1.03 Petaflops. (#11 Top500)

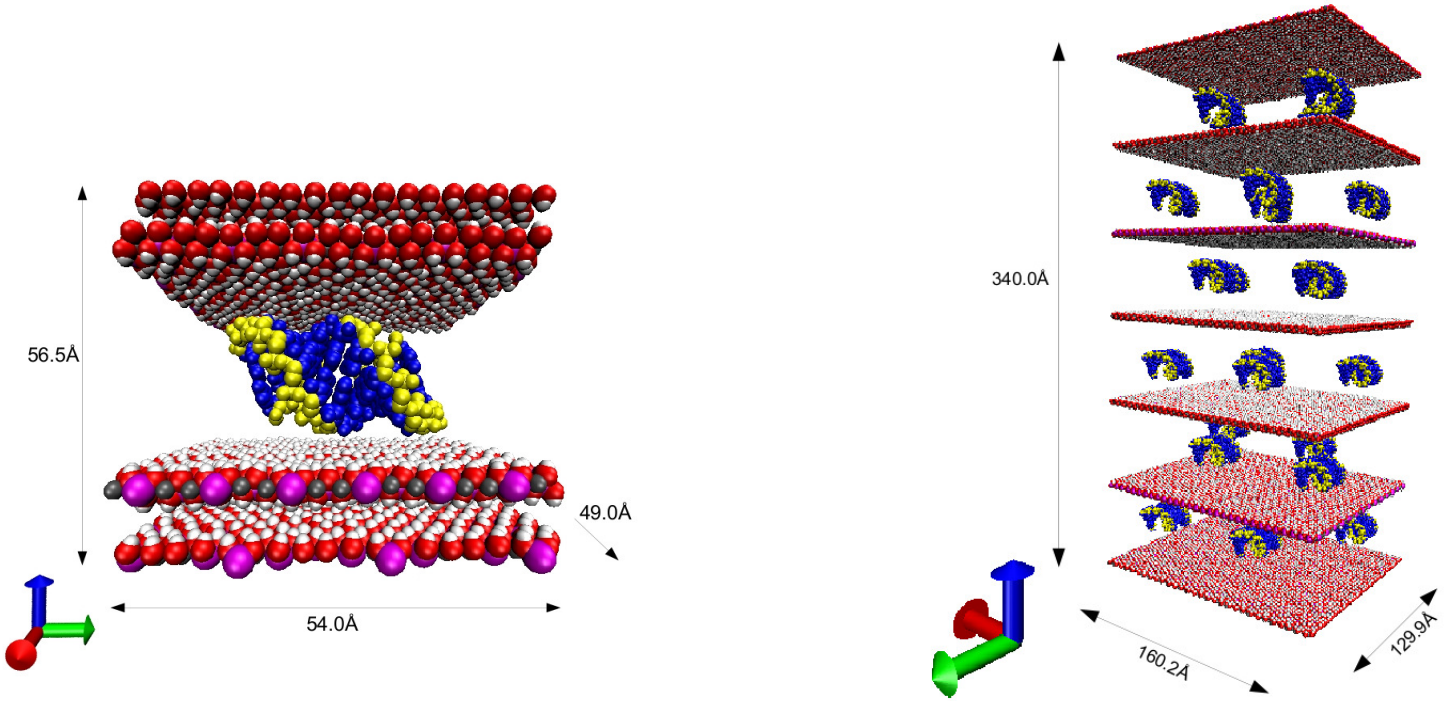
Jugene (*Jülich Blue Gene*), at Forschungszentrum Jülich
IBM Blue Gene machine with 294,912 cores. A peak performance of 1.00 Petaflops (#12 Top500).



Kraken, NICS

Jugene, FZJ-JSC

Simulation of LDH-clay nucleic acid systems

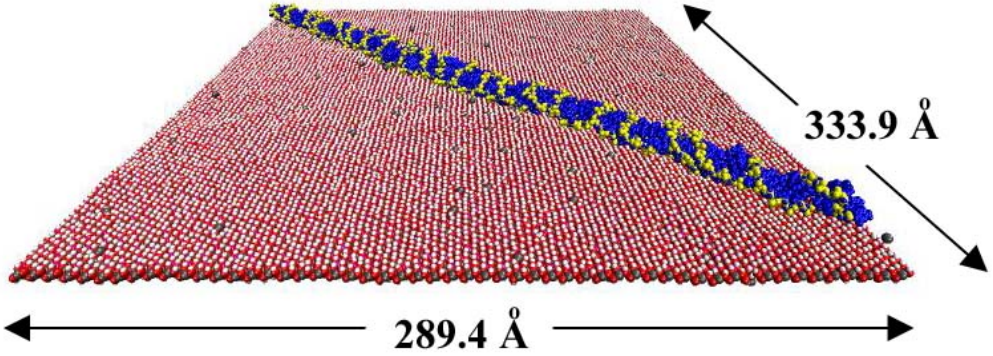


The layered double hydroxide (LDH) which forms the basis for the mineral models employed in this study has unit formula $[Mg_2Al(OH)_6]\cdot H_2O.Cl$. The LDH models are replicated from a unit cell with dimensions 16.34Å x 18.82Å x 25.34Å, obtained by the refinement of powder X-ray diffraction data on hydrotalcite using Rietveld methods.

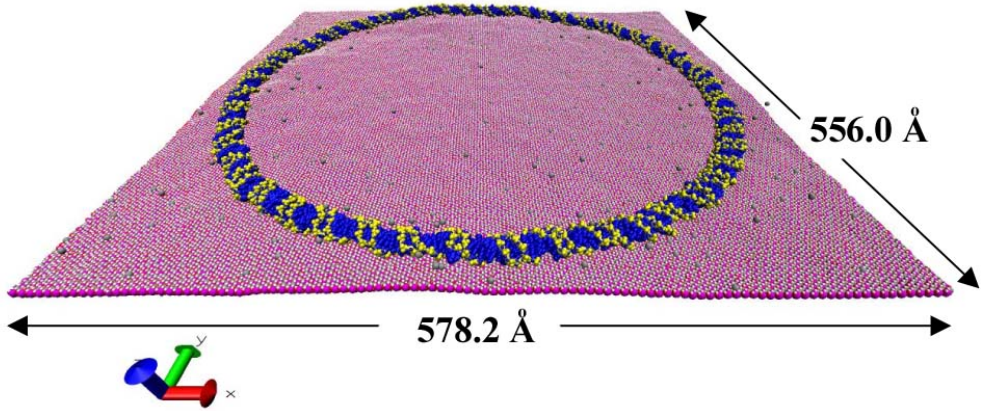
Simulation of LDH-clay DNA systems

Stability within MgAl-LDH at ambient conditions

- Very little movement in DNA within 1ns of simulation
- DNA compressed
  - Number of base pairs < DNA in bulk water
- DNA heavily influenced by movement of LDH sheets
  - Thermal undulations of LDH



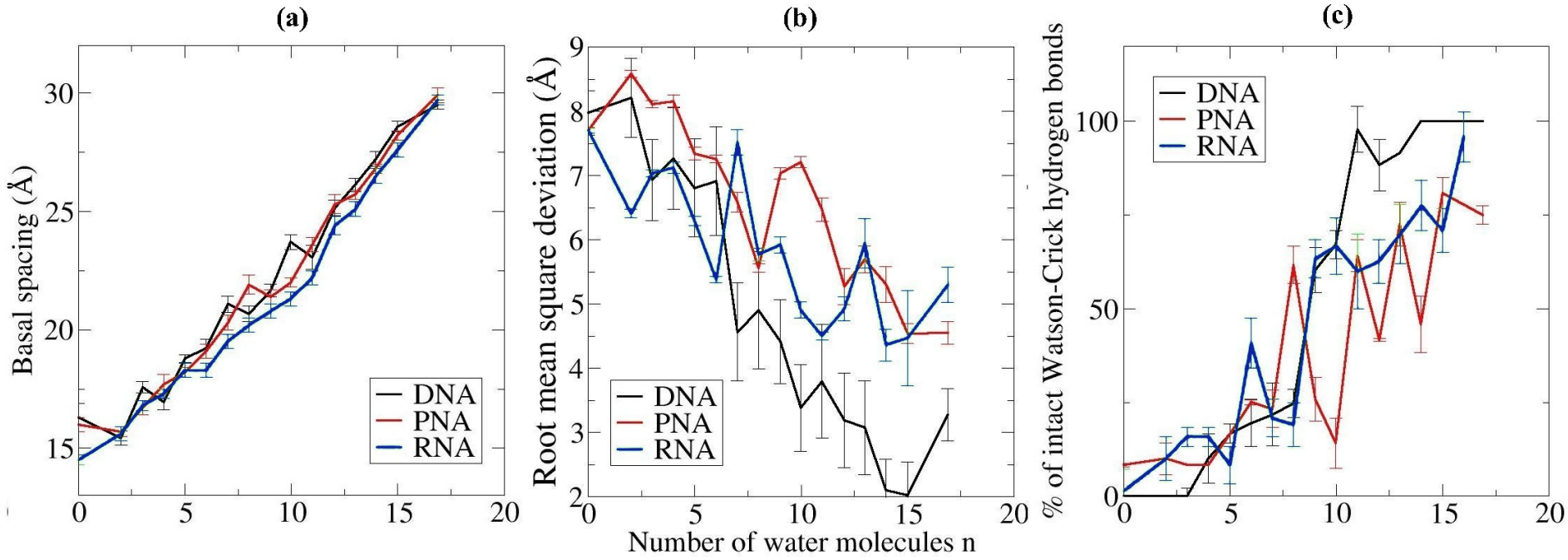
344,959 atoms, 108bp



1,157,038 atoms, 480bp

Thyveetil, M.-A. Greenwell H.C. and Coveney, P.V.(2008) *J. Am. Chem. Soc.* 14: 4742-4656

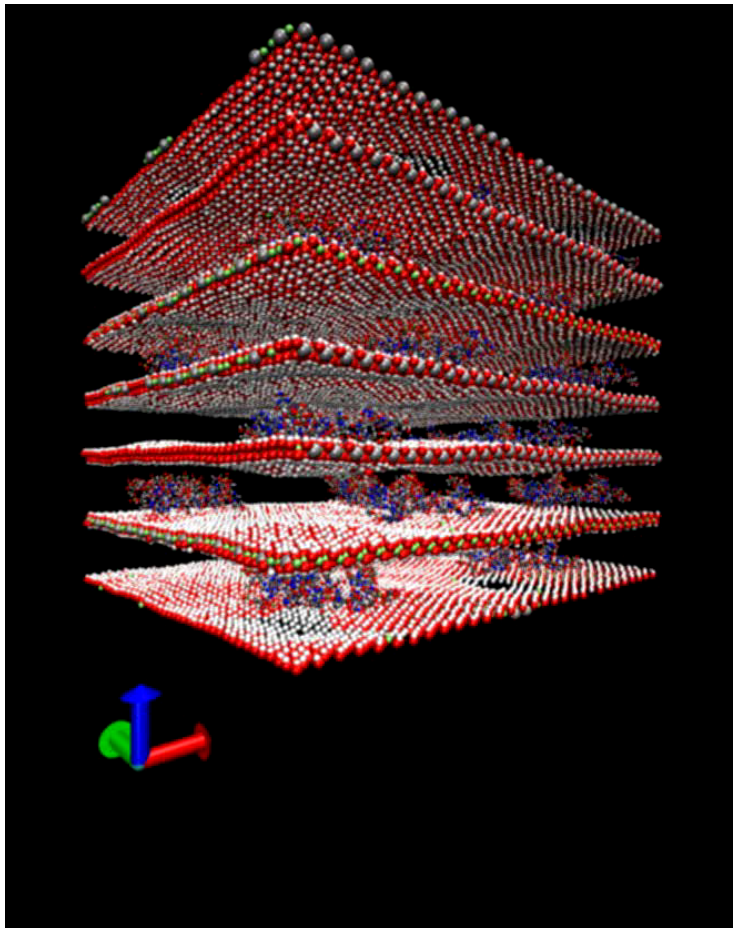
Properties of nucleic acids within LDH-clay



- RNA double strands are the most easily compressed.
- Intercalated PNA duplex strands have the largest RMSD values indicating that the base-pairing in this system is significantly disrupted.
- Intercalated DNA conserves the most Watson-Crick base pairs.

Swadling, J. Greenwell H.C. and Coveney, P.V. (2011) Stability of Free and Mineral-Protected Nucleic Acids: Implications for the RNA World. *Preprint.*

Simulation of LDH-clay nucleic acid systems



RNA intercalated within an Mg_2Al LDH

- 300 K, 1 atm.
- NPT ensemble
- 0.5 fs timestep.
- 300 k atoms, 30 ns.

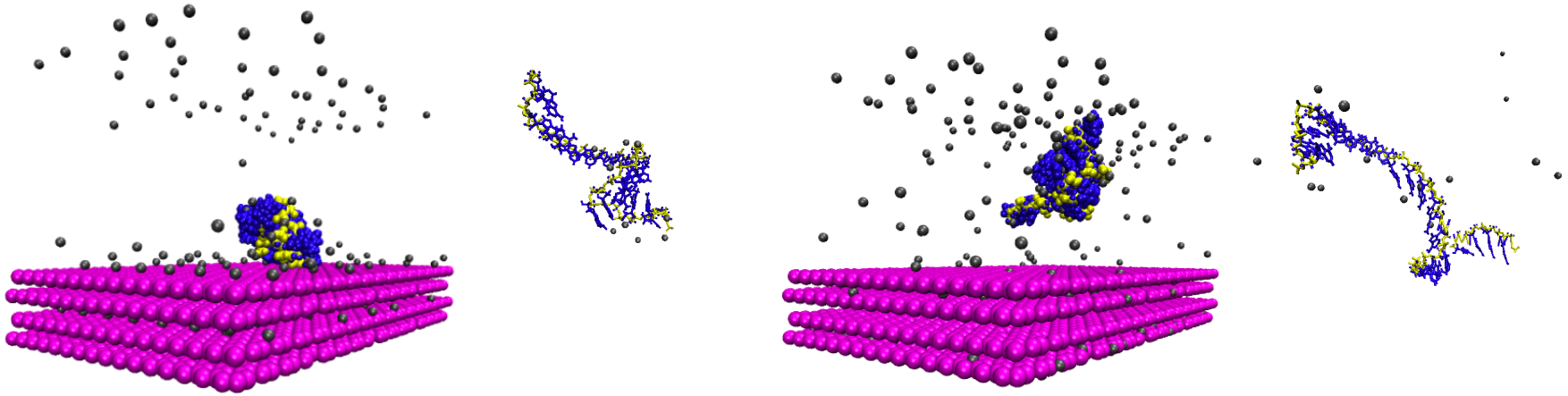
Simulation shows:

- LDH corrugates around RNA
- RNA stabilized by LDH
- RNA restricted in motion and concentrated.
- Compartmentalisation of biomolecules.

J. B. Swadling, P. V. Coveney, & H. C. Greenwell. *Geochim. Cosmochim. Acta*, 83:360-378, 2012.

Simulation of smectite-clay/RNA systems

Single-stranded RNA with an exposed aqueous montmorillonite clay surface.

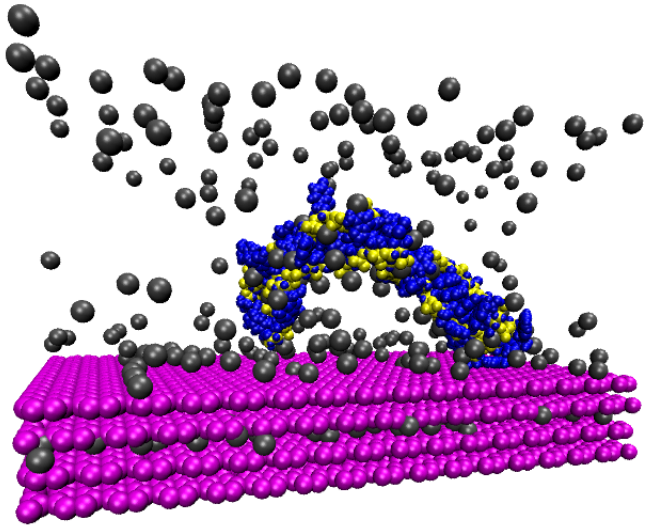


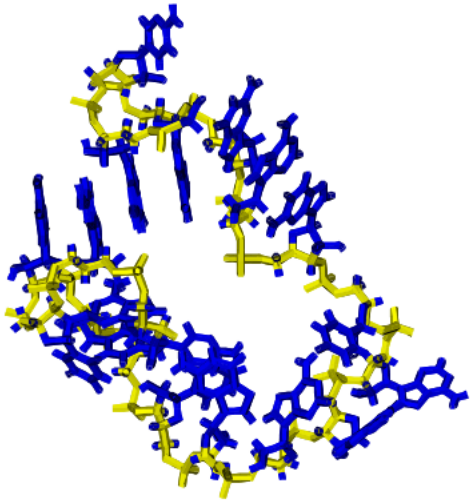
Over timescales of only a few nanoseconds, specific RNA sequences fold to characteristic secondary structural motifs, which do not form in the corresponding bulk water simulations.

Swadling, J. Greenwell, H.C. and Coveney, P.V. (2010) Clay Minerals Mediate Folding and Regioselective Interactions of RNA: A Large-Scale Atomistic Simulation Study. *J. Am. Chem. Soc*. 132(39), 13750-13764.

Simulation of smectite-clay/RNA systems

Our simulations 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.





A stem-loop secondary structural motif of RNA formed on the clay surface, stabilized through base stacking.

Swadling, J. Greenwell, H.C. and Coveney, P.V. (2010) Clay Minerals Mediate Folding and Regioselective Interactions of RNA: A Large-Scale Atomistic Simulation Study. *J. Am. Chem. Soc*. 132(39), 13750-13764.

Conclusions

Theory, modeling & simulation provide a powerful set of concepts and tools to assist in origins of life research.

- **Theory of complex non-linear chemical kinetic schemes can provide explanations for**
  - origin of self-replicating systems, such as RNA
  - origin of chirality & homochirality in polymers and crystals

- **Molecular simulation**
  - allows for study in exquisite detail of interactions, catalysis, conformations and folding of candidate molecular entities (nucleic acids, proteins, clays, etc.)
  - provides a way of probing extreme conditions without experimental problems

Acknowledgments

All of the research reported in this talk, and more, is published within a *Chemical Society Reviews* special issue, edited by Professor Piero Ugliengo.

Funding and Support