

SELF-PROGRAMMABLE, SELF-ASSEMBLING TWO-DIMENSIONAL GENETIC MATTER

STEPHEN J. SOWERBY^{1,*}, PETER A. STOCKWELL¹, WOLFGANG M. HECKL²
and GEORGE B. PETERSEN¹

¹ *Department of Biochemistry and Centre for Gene Research, University of Otago, P.O. Box 56, Dunedin 9001, New Zealand*

² *Institute für Kristallographie, Ludwig Maximilians Universität, 80333 München, Germany*

(* Author for correspondence, e-mail: sowerby@sanger.otago.ac.nz)

(Received 14 June, 1999)

Abstract. Putative two-dimensional coding systems can be constructed from aqueous solutions of purine and pyrimidine nucleic acid bases evaporated at moderate temperatures on the surfaces of inorganic solids. The resultant structures are monolayers which are formed spontaneously by molecular self-assembly and they have been observed with molecular resolution by scanning tunnelling microscopy (STM). When formed from solutions of a single base, the monolayers of adenine and uracil have crystalline characteristics and the STM images can be interpreted in terms of the geometrical placement of planar arranged molecules that interact laterally by intermolecular hydrogen bonding. When formed from solutions containing a mixture of adenine and uracil, the monolayers have aperiodic structures. Small crystalline domains within these monolayers can be interpreted in terms of the single phase configurations of the molecules and the remaining aperiodic structures can presumably be interpreted, geometrically, in terms of the 21 theoretically possible adenine-adenine, uracil-uracil and adenine-uracil hydrogen bonding interactions. We propose that combinatorial arrangements of planar arranged purine and pyrimidine bases could provide the necessary complexity to act as a primitive genetic mechanism and may have relevance to the origin of life.

1. Introduction

In modern biological systems, the non-periodic sequence of purine and pyrimidine bases in nucleic acid molecules encodes the order of amino acids in linear peptide polymers. Following chemical linkage of the amino acids, the peptides spontaneously form higher-ordered three-dimensional structures (proteins) that can function with catalytic capabilities. The use of proteins as catalytic agents is a hallmark of modern biochemical systems and the primary order of the amino acids in the peptides is the only determinant of protein structure and function (Anfinsen, 1973). It is clear that a key step in the origin of life was the development of a mechanism to encode the primary order of the amino acids in proteins.

The prebiotic availability of organic molecules is supported by abiotic chemical synthesis (Baudisch, 1913; Löb, 1913) and chemical simulation experiments of the primitive earth (Miller, 1953; Ferris and Hagan, 1984). The predominant products of such studies are a selection of the biological amino acids (Miller, 1953).



Origins of Life and Evolution of the Biosphere **30**: 81–99, 2000.
© 2000 Kluwer Academic Publishers. Printed in the Netherlands.

Similarly, some of the purine and pyrimidine bases found in nucleic acids can be prepared *ex vivo* and are synthesised from the same cyano intermediates that give rise to the amino acids (Ferris and Hagan, 1984). This suggests that conditions may have prevailed for the simultaneous prebiotic availability of both amino acids and bases. The potential routes to the sugar components of nucleic acids, however, are not consistent with these chemistries (Schwartz and de Graaf, 1993) and the *de novo* synthesis of nucleic acid polymers would seem extremely implausible under these circumstances (Cairns-Smith, 1988; Joyce, 1989). What, then, constituted the earliest coded protein synthesis mechanism? Could it have been constructed utilising just amino acids and bases alone?

We propose that free purine and pyrimidine bases, alone, adsorbed to a suitable surface, could have constituted a primitive coding template for the construction of catalytic proteins (Sowerby, 1995; Sowerby *et al.*, 1996). This hypothesis is based on the demonstration by ourselves and others that purine and pyrimidine bases can spontaneously self-assemble into stable two-dimensional monolayer structures at the interface between water and the surfaces of crystalline inorganic solids (Allen *et al.*, 1991, 1992; Heckl *et al.*, 1991; Srinivasan *et al.*, 1991; Srinivasan and Murphy, 1992; Heckl, 1993; Srinivasan and Gopalan, 1993; Tao *et al.*, 1993; Bolland and Ratner, 1994; Tao and Shi, 1994a, b, c, d; Poler *et al.*, 1995; Sowerby, 1995; Sowerby *et al.*, 1996, 1998a, b, c; Sowerby and Heckl, 1998; Sowerby and Petersen, 1997, 1999).

The suggestion that inorganic mineral surfaces might have been involved in the origin of life is not new (Bernal, 1951). The formation of peptide bonds between non-coded amino acids adsorbed on mineral surfaces has been well demonstrated (Brack, 1993; Bujdak and Rode, 1996; Hill *et al.*, 1998; Orgel, 1998) and it has been proposed that imperfections in the structures of inorganic scaffolds, such as clay, could have constituted a primordial genetic coding system (Cairns-Smith, 1982, 1988). In our monolayer model, however, the inorganic surface has no informational content. It is the spatial arrangement of the bases that facilitates information storage and the planar arranged bases specify the primary order of the amino acids in proteins. In nucleic acid polymers, the information is specified by the spatial arrangement of the purine and pyrimidine bases which are fixed relative to one another by attachment to the sugars in the sugar-phosphate scaffold. In the monolayers, the bases are also fixed relative to one another, but in matrices arrayed on a flat two-dimensional scaffold.

The purine and pyrimidine bases are planar heterocyclic molecules which contain both proton acceptor and proton donor substituents and hydrogen bonding interactions between them facilitates molecular recognition during biological information processing. On flat, uncharged surfaces the bases are planar-arranged like jigsaw puzzle pieces on a table and hydrogen bonds between the bases can be likened to the interlocking features of the jigsaw puzzle which specify the matching rules between adjacent pieces. Monolayers of the bases were first observed at a mercury-water interface using electrochemical methods (Vetterl, 1965,

1966) and the literature on these and subsequent electrochemical studies has been reviewed (De Levie, 1988; De Levie and Wandlowski, 1994). But the self-assembly of the monolayer structures on inorganic surfaces was not studied until the advent of scanning tunnelling microscopy (STM, Binnig *et al.*, 1982). STM relies on quantum electron tunnelling between the terminal atoms of a sharp metal probe raster scanned over the surface atoms of a conducting or semiconducting solid. The output of STM experiments is an image of the surface features that can be likened to a coin rubbing with atomic resolution. Because STM is not a surface averaging technique, it is ideally suited to the high resolution analysis of periodic and non-periodic surface structures with atomic level detail. A review of the STM studies of monolayers of bases has been published (Sowerby and Heckl, 1998).

The purine and pyrimidine bases are sufficiently soluble in water (Dawson *et al.*, 1986) to allow direct application of aqueous solutions of the bases to van der Waals surfaces like the basal planes of graphite and MoS₂. Although graphite and MoS₂ might not be ubiquitous prebiotic compounds, van der Waals surfaces are organophilic and the prebiotic availability of surfaces of this nature suggests that the interactions of organic molecules with them could have played a role in the origin of life (Smith, 1998). Because graphite and MoS₂ {0001} surfaces have been well characterised by STM: Graphite {0001} (Park and Quate, 1986) and MoS₂ {0001} (Stupian and Leung, 1987; Weimer *et al.*, 1988), they are ideal model surfaces to study self-assembly phenomena (Sowerby and Petersen, 1999).

Close packed monolayers formed on the surfaces of inorganic compounds are particularly well suited to examination by STM because their structures are resistant to the forces of the scanning STM probe. On graphite, the STM images of the monolayer adsorbates are highly convoluted and the size of the image features corresponds to submolecular detail. Interpretation of the images is complicated by the substrate-dependent modulation of the image. On MoS₂, however, the individual molecules can be identified discretely and the interpretation becomes a problem of geometry (Heckl *et al.*, 1991; Heckl, 1993; Sowerby *et al.*, 1996, 1998c; Sowerby and Petersen, 1997, 1999). Monolayer formation occurs spontaneously on the solid surface from aqueous solutions containing the dissolved species and is accelerated by heat (Allen *et al.*, 1991; Heckl *et al.*, 1991) or salt (Tao and Shi, 1994b) and is thus a plausible mechanism within localised environments of the prebiotic earth. Although the prebiotic availability of homogeneous preparations of bases seems unlikely, preliminary investigations have focused on the contrived structures formed from artificially pure preparations on model surfaces. This has been to develop an understanding of the self-assembly principles and the STM image contrast of this class of molecules.

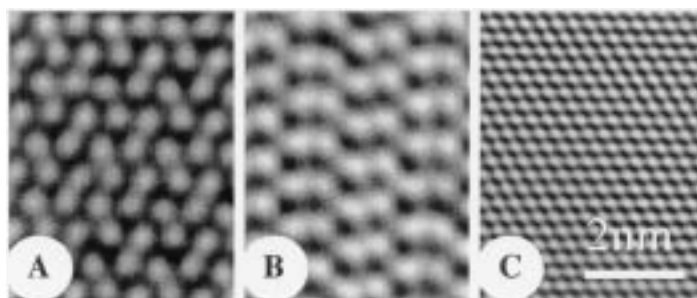


Figure 1. STM images of monolayers of uracil (A) and adenine (B) self-assembled on the surface of MoS₂. The STM image of the MoS₂ surface underlying the uracil monolayer (C) shows the local density of states of the uppermost sulfur atoms of the cleaved {0001} MoS₂ surface and are separated by a distance of 0.316 nm. Comparison of the adsorbate images (A,B) with those of the directly underlying substrate (C) shows that the dimensions of the individual features agree well with the lateral dimensions of planar arranged adenine and uracil molecules. The STM images of the monolayers (A,B) constitutes molecular resolution.

2. Adenine and Uracil Monolayers

The purine base, adenine, and the pyrimidine base, uracil, have previously been studied as single phase monolayer systems formed on the surfaces of graphite and MoS₂ (Allen *et al.*, 1991; Srinivasan *et al.*, 1991; Srinivasan and Murphy, 1992; Heckl, 1993; Srinivasan and Gopalan, 1993; Tao and Shi, 1994b; Freund *et al.*, 1997; Sowerby and Petersen, 1997; Edelwirth *et al.*, 1998; Sowerby *et al.*, 1996, 1998a, b, c; Sowerby and Heckl, 1998). STM images of monolayers of adenine and uracil on the surface of MoS₂ are shown in Figure 1.

The interaction of the molecules with the surface is physisorption-predominantly weak, non-covalent interactions of the van der Waals type. Laterally, the molecules are stabilised within the monolayer structure by van der Waals packing constraints and cyclic hydrogen bonds. π -Bond cooperativity, which contributes to the stability of cyclic hydrogen bonds between complementary base pairs in nucleic acids, requires that adjacent, hydrogen bonding functional groups are linked by bonds with π -electron character (Figure 2) (Jeffrey and Saenger, 1994). In these interactions, electron delocalisation results in the smearing of electrons over the base-base dimers. This electronic effect is clearly evident in the STM image of the uracil monolayer (Figure 1A) where dimers stabilised by cyclic hydrogen bonds form a herringbone pattern. The function of the scan angle on the STM image contrast of adenine monolayers shows how the STM tip can cause electron delocalisation in the monolayer. This can complicate the image interpretation in these systems, where extended π -electron systems are formed through the electronic coupling of aromatic molecules linked by cyclic hydrogen bonds (Sowerby *et al.*, 1998c).

The hydrogen bonding interactions specify the molecular configuration of the monolayers and, as in nucleic acids, are a mechanism for molecular recognition between adjacent planar-arranged bases which suggests the possibility of mono-

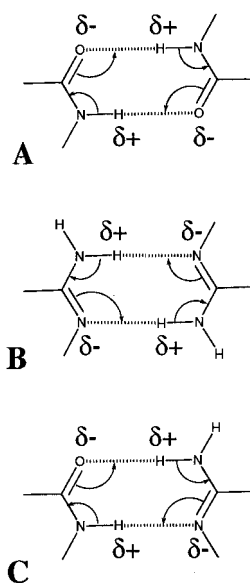


Figure 2. Cyclic hydrogen bonding of the types proposed for the interactions between adenine and uracil molecules. The arrows indicate resonance assisted electronic contributions to the cyclic hydrogen bond structure.

layers of mixed purine and pyrimidine composition. Examination of the available crystal data of purine and pyrimidine derivatives indicates that mixed purine-pyrimidine (Hoogsteen, 1963; Mathews and Rich, 1964; O'Brien, 1967), purine-purine (Sakore and Sobell, 1969) and pyrimidine-pyrimidine (Voet and Rich, 1969) complexes form readily in the solid state and these, also, are stabilised by intermolecular hydrogen bonding. Preliminary STM evidence supported mixed purine and pyrimidine monolayer formation between guanine and uracil but the aperiodic structures proved unstable against the scanning action of the STM probe and were swept away (Sowerby and Heckl, 1998).

Jeffrey and Saenger (1994) describe purine and pyrimidine homo and heteropairing with the notation ABM^N , where A and B are single letter codes for bases, M is a sequential numerical identifier and N is the number of hydrogen bonds in the interaction. Using this notation, a mixture of adenine and uracil molecules can theoretically form 21 dimer combinations that are stabilised by two adjacent hydrogen bonds to form cyclic hydrogen bonded structures (Figure 3). Single hydrogen bonding geometries, of the type seen in the uracil monolayer (Figure 4A), are also possible but because interactions of the cyclic type would be favoured preferentially (Jeffrey and Saenger, 1994) they are the only type that we have considered here.

Of the possible dimeric structures shown in Figure 3, only some have been observed experimentally (Jeffrey and Saenger, 1994). In the solid state, homo-base pairs in a centrosymmetric configuration are favoured because the purines

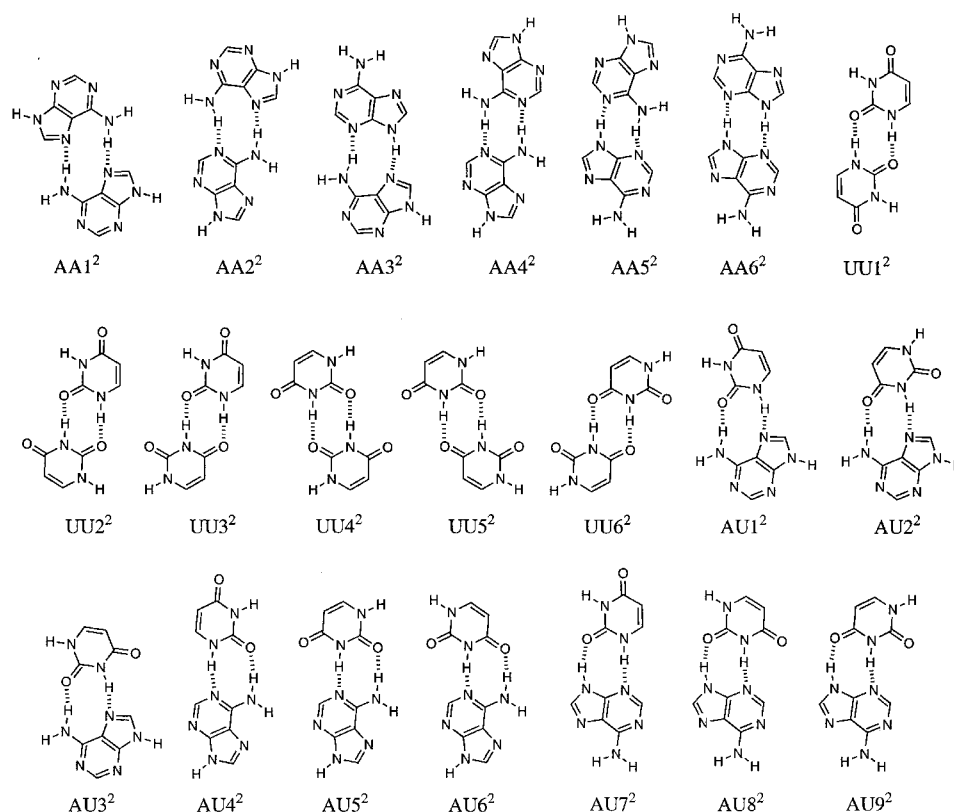


Figure 3. The theoretically possible homo- and hetero dimeric forms of adenine-adenine, uracil-uracil and adenine-uracil interactions which are stabilised by cyclic hydrogen bonding. Adenine and uracil have pKa values of <1, 4.1, 8.9 and 0.5, 9.5, >13, respectively (Dawson *et al.*, 1986) and suggests that the hydrogen bonding configurations between adenine and uracil molecules will be consistent within the pH range 4.1–8.9. Adenine and uracil have the ability undergo amino-imino and keto-enol tautomerism, respectively, but at neutral pH the bases are not ionised and the equilibria favour the amino and keto canonical forms (Jeffrey and Saenger, 1994).

and pyrimidines have large dipole moments which are cancelled in an antiparallel configuration. This results in cancellation of the electric field over the crystal volume (Jeffrey and Saenger, 1994). In crystalline uracil (Parry, 1953), the UU6² centrosymmetric dimer packs into flat sheets stabilised by singular, non-cyclic hydrogen bonding interactions between adjacent dimers. This configuration is reproduced in the monolayer structure on MoS₂ (Figure 4A) and graphite surfaces (Sowerby and Petersen, 1997). All of the other UU configurations have also been observed, either in the solid state or within nucleic acid complexes (Jeffrey and Saenger, 1994). Similarly, the adenine monolayer structure (Figure 4B) is composed of homodimers of the AA6² type with adjacent dimers stabilised by non centrosymmetric AA2² cyclic interactions (Edelwirth *et al.*, 1998; Sowerby *et al.*, 1998b, c). This is evidence that cyclic hydrogen bonding interactions are not mu-

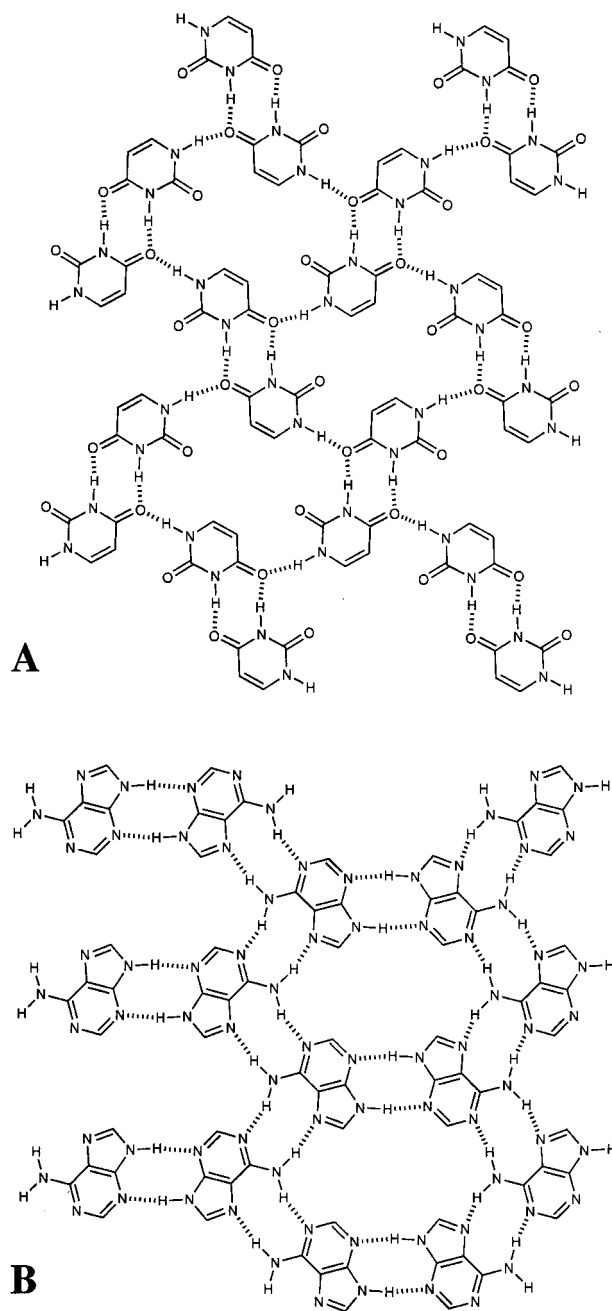


Figure 4. Models of the molecular arrangement of uracil (A) and adenine (B) monolayer configurations determined previously by molecular modelling and energy minimisation (Edelwirth *et al.*, 1998; Sowerby *et al.*, 1998a, b, c). Both adenine and uracil molecules are in planar arrangement on the surface and the models can be directly related to the STM images (Figures 1A, B). Hydrogen bonds between the molecules are indicated by the hashed lines.

tually exclusive. Each molecule can participate in more than one cyclic interaction. The solid state structure of pure adenine has not been determined but the AA1² and AA4² centrosymmetrical base pair interactions have been observed in the protonated adenine·HCl structure (Kistenmacher and Shigematsu, 1974) and in N9 substituted complexes (Jeffrey and Saenger, 1994) respectively. Heterobase interactions between adenine and uracil have been observed only for the reversed Hoogsteen AU2², Hoogsteen AU3², reversed Watson-Crick AU5² and Watson-Crick AU6² base pairs (Jeffrey and Saenger, 1994). However, in these structural complexes, the purine N9 and pyrimidine N1 proton donor sites were substituted as they are in nucleic acids. Interactions involving these sites in the free bases have only been observed in the hydrogen bonded monolayers.

In this communication, we present evidence obtained by STM for the self-assembly of monolayers formed from mixtures of adenine and uracil bases on the surface of MoS₂. Our aim is to demonstrate that complex aperiodic structures can emerge from simple putative prebiotic molecules through their spontaneous self-assembly on the surfaces of inorganic solids. We propose that these self-construction mechanisms could have led to the genesis of primordial biological information.

3. Methods

The monolayer structures were prepared by thermally-driven evaporation following application of aqueous solutions of the bases (10 μ L, neutral pH) to the heated (65 °C) {0001} surfaces of freshly cleaved naturally occurring MoS₂. The single component systems were applied as saturated aqueous solutions of the bases. The mixed component system was applied as a dilute aqueous solution containing uracil, 194 μ M, and adenine, 31 μ M, which corresponds to a molecular ratio of approximately 6:1.

All STM images presented here were obtained with a Nanoscope II (Digital Instruments Inc., Santa Barbara CA, U.S.A.) operated under ambient conditions using electrochemically etched (2 M KOH, 10 V AC) polycrystalline tungsten probes. The measurements were made at bias voltages of ± 5 to ± 1200 mV with the tunnel current set to values between 100 pA and 3 nA. Images of the adsorbate structure were calibrated internally using images of the underlying substrate obtained as consecutive image frames of the adsorbate and of the directly underlying substrate by deliberately rupturing the adsorbate following reduction of the bias voltage. Nonlinearities in the piezoelectric scanner and the scan generating mechanism were corrected by assuming a perfect three-fold symmetry for the images of the underlying substrate and the known crystallographic dimensions of MoS₂ which has an in-plane lattice constant of 0.316 nm. The geometric corrections were performed using the software 'Image SXM' a spin-off of the public domain 'NIH Image' program (Barrett, 1998). Fourier space filtering of the STM images was

done by selecting and retaining the most intense reciprocal space reflections using the thresholding and look-up-table (LUT) tools in 'Image SXM', thereby removing spurious noise signals from the real-space image.

Molecular mechanics energy minimisation and energy calculations were performed with the program 'SPARTAN 5.1.1' (Wavefunction Inc. Irvine CA, U.S.A.) using the Merck Molecular Force Field accessible through 'SPARTAN'. The stabilisation energies of hydrogen bonded dimers were determined by calculating the energy difference between the dimer and the sum of the energies of the two separated molecules (Sponer *et al.*, 1996). The deviations of the dimers from planarity were determined from the energy minimised models by measuring the angle between the planes of the two molecules. The molecular planes were defined by the aromatic ring component of each of the purine and pyrimidine bases.

4. Results and Discussion

4.1. STM OF PURINE AND PYRIMIDINE MONOLAYERS

In order to obtain images of mixed composition monolayers (Figure 5), trial and error experiments using biases between the molar ratios of the two bases in the applied solution were required. For uracil and adenine mixtures, a molar ratio of approximately 6:1 gave results suggesting aperiodic monolayer formation (Figure 5). We did not observe formation of aperiodic structures with ratios greater than 12:1 or below 6:1. Binary mixtures outside these ranges gave periodic lattices consistent with pure monolayers of the predominant base component.

In the STM images showing aperiodic adsorbate formation (Figures 5A, B), local regions of periodic structure can be observed but overall regularity is absent, which suggests the possibility of mixed base composition. The domains of local crystallinity have lattice periodicities and packing configurations consistent with that of pure adenine and the unit meshes of two enantiomorphic forms of the adenine monolayer structure are simultaneously distinguishable (Figures 5C, D). We had previously observed that adenine, although achiral, could form two mirror symmetric structures on the surface of MoS₂ and have proposed that this could have played some role in prebiological chiral symmetry breaking (Sowerby *et al.*, 1996, 1998c).

The regions of pure adenine periodicity are useful from an analytical point of view because their well-characterised molecular dimensions and packing structure serve as an internal calibration. In these regions, the adenine molecules are clearly resolved and, in both the unfiltered (Figure 5A) and filtered images (Figure 5B), the molecules have a triangular appearance. Although the molecules lying near the edges of the cluster are clearly resolvable, these features are less well defined, presumably because molecules not within the periodic structures have more opportunity for lateral motion induced by the scanning action of the STM probe. The

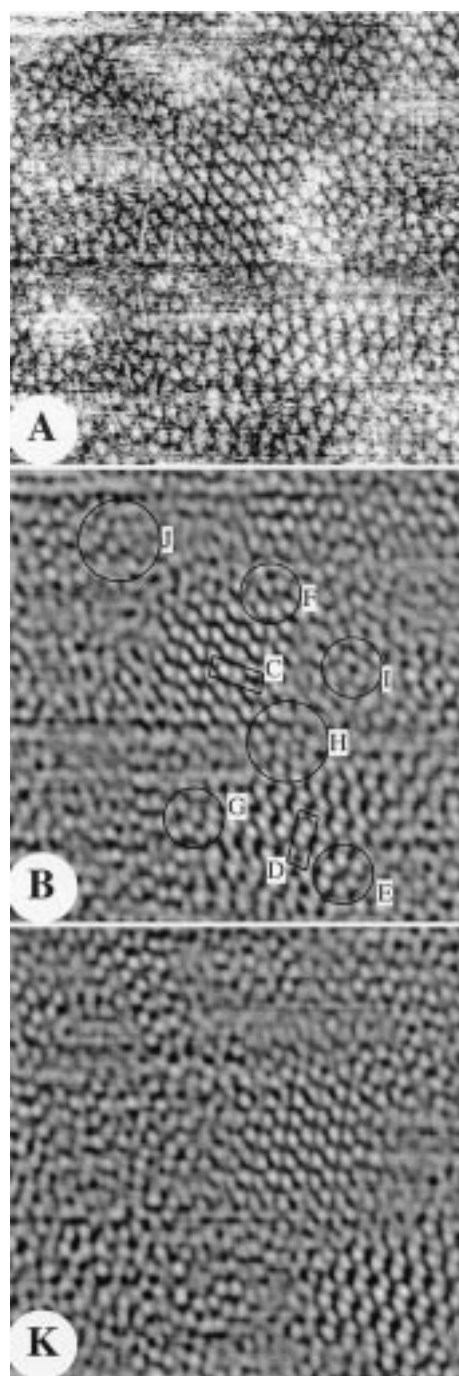


Figure 5. Raw STM image (A) and filtered STM image (B,K) of a region of aperiodic adsorbate structure formed on the surface of MoS_2 from a mixture of adenine and uracil molecules (scan size 17×17 nm).

resolution of the adenine molecules within their periodic lattice arrangement does, however, indicate that the image contrast features attributed to adenine species are true molecular components in the STM images and are not artefactual. This gives confidence in the assessment of aperiodicity within non-crystalline regions of the images.

The edges of the adenine clusters provide the first evidence of aperiodic structure formation (Figures 5E, F, G, H). In these regions, discrepancies in the lattice packing are easily distinguished and suggest adenine-adenine or adenine-uracil interactions. Discrimination between adenine and uracil molecules, however, is not straightforward. Even though the purine molecules have a triangular appearance within the cluster of the periodic structures, the shape definition wanes towards the edges. The van der Waals volume and surface area for uracil (118.47 \AA^3 , 137.74 \AA^2) are only 82 and 86% of the adenine values, respectively (145.03 \AA^3 , 160.43 \AA^2) and the level of noise in the STM images makes molecular identification based on shape or size difficult. Consequently, we have been unable to make a reliable assessment of the ratios of adsorbed molecular species. The images also display regions that are clearly composed of discrete molecular structures, but which exhibit no long or short range order (Figure 5J). These too, are suggestive of homo, or hetero-base aperiodicity. In the preceding image frame (Figure 5K) only regions that show some order are identifiable in the subsequent Figure 5B frame suggesting that the energy provided by the scanning action of the STM probe may have re-ordered aperiodic fragments of the lattice.

Although our primary aim is to demonstrate that the coding components of polymeric nucleic acids can self-assemble into two-dimensional matrices of mixed composition, it is difficult at this stage to give an unambiguous interpretation of the STM images of mixed composition. This is because interpretation of the aperiodic structures that we have observed is complicated by the 21 adenine and uracil interactions that must be considered. However, where regions of aperiodicity give way to well characterised domains of known composition and molecular configuration, the interpretation can be related to the single component studies and this provides an intuitive starting point for determination of the allowable base-base interactions and two-dimensional arrangements.

Parallel to self-assembly studies of the purines and pyrimidines at the solid-fluid interface, are ultra high vacuum (UHV) studies of monolayers prepared by sublimation techniques (Tanaka and Kawai, 1995; Tanaka *et al.*, 1996a, b; Freund *et al.*, 1997; Furukawa *et al.*, 1997; Kawai *et al.*, 1997; Nakagawa *et al.*, 1997). Although the preparation procedure is quite different, adenine monolayers prepared on graphite surfaces from aqueous solvents were determined to be identical to those prepared by sublimation in UHV (Sowerby *et al.*, 1998b). Using sublimation techniques, well characterised submonolayer deposition of the purine and pyrimidine bases has been achieved (Tanaka *et al.*, 1996a, 1996b; Freund *et al.*, 1997; Kawai *et al.*, 1997; Nakagawa *et al.*, 1997) and suggests an experimental route to well-

controlled analyses of the interactions between purine and pyrimidine bases on inorganic surfaces.

4.2. COMPUTER SIMULATION

A first order approximation of the possible configurations of homo- and hetero-base pairs can be achieved by examining the stabilisation energies of the dimeric combinations. These are determined using computer modelling and by calculating the energy difference between the dimer and the sum of the energies of the two separated molecules given by: $\Delta E = (E^{AB}) - (E^A + E^B)$ (Sponer *et al.*, 1996). Calculations have been made for various purine and pyrimidine homo- and hetero-dimers at the molecular mechanics, semi-empirical and *ab initio* levels of calculation (Sponer *et al.*, 1996) and provide a theoretical route to study the determinants which specify adjacent base-base interactions.

We have determined the energy minimised structures and stabilisation energies for the 21 possible adenine and uracil homo and hetero-dimer combinations, shown in Figure 3, at the molecular mechanics level of calculation using the Merck Molecular Force Field (MMFF, Table I). Also measured from the energy minimised structures are the deviations of the dimers from planarity.

Monolayers formed from adenine on graphite and MoS₂ used the configuration of centrosymmetric dimer, AA6², which had the greatest molecular mechanics level stabilisation energy calculated using the Dreiding II Force Field (Edelwirth *et al.*, 1998; Sowerby *et al.*, 1998b). Similarly, the AA6² dimer has the greatest stabilisation energy of the centrosymmetric configurations (Figure 3, AA1², AA4², AA6²) determined by the MMFF. The monolayers formed from the purine base, xanthine, was resolved in a similar way to adenine but also showed that cooperative effects may lead to stabilisation where cyclic hydrogen bond motifs become linked (Sowerby and Petersen, 1999). The uracil monolayer (Figure 4B) is constructed from the UU6² centrosymmetric dimer models determined for the solid state by X-ray crystallography (Sowerby and Petersen, 1997; Sowerby *et al.*, 1998a) but this configuration does not have the greatest stabilisation energy of the centrosymmetric dimers (Figure 3, UU1², UU4², UU6²) determined by MMFF (Table I).

It also seems sensible to consider that dimers which deviate considerably from planarity would not exhibit favourable interactions when restricted in-plane by adsorption on a flat surface, compared with dimers which do not have significant deviations from planarity. This is because the native state interaction would have to overcome a deformation energy in order to achieve the planar configuration. For the adenine monolayers, the AA6² and AA2² dimer motifs from which the monolayers are constructed do deviate from planarity, 0.6 and 37.7°, respectively. However, these deviations were not seen in the fully minimised models of the monolayers (Edelwirth *et al.*, 1998; Sowerby *et al.*, 1998b), which suggests that the MMFF may over estimate the degree of pyrimidalisation contributed by the lone pair electrons on the amino group of adenine. This effect is reduced significantly at the higher

TABLE I

The calculated stabilisation energies and deviations from planarity from energy minimised models of cyclic hydrogen bonded adenine and uracil homo- and hetero-dimers

Dimer ABM ^N	Stabilisation energy (kCal mol ⁻¹)	Deviation from planarity (°)
AA1 ²	-8.41	69.1
AA2 ²	-8.42	37.7
AA3 ²	-11.29	35.7
AA4 ²	-7.00	0.6
AA5 ²	-9.92	41.9
AA6 ²	-10.40	0.6
UU1 ²	-13.80	0.0
UU2 ²	-12.76	0.0
UU3 ²	-12.96	0.0
UU4 ²	-12.38	0.0
UU5 ²	-12.26	0.0
UU6 ²	-12.38	0.0
AU1 ²	-11.01	34.6
AU2 ²	-12.58	23.2
AU3 ²	-12.62	22.6
AU4 ²	-9.78	39.8
AU5 ²	-11.27	17.5
AU6 ²	-11.31	17.9
AU7 ²	-12.61	3.0
AU8 ²	-13.23	0.3
AU9 ²	-13.45	1.3

level *ab initio* calculations of the Hartree-Fock type (Sponer *et al.*, 1996).

When considering the hetero-base interactions, it may be possible to exclude some of the 21 possible base-base interactions on geometry and energy arguments. The deviations from planarity of the AA3²-AA5² and AU1²-4² are significant according to the MMFF (Table I). However, the planar arrangement and the calculated stabilisation energies for the AU5²-9² and the remaining homo-base dimers would not exclude these interactions based on geometry or energy arguments alone. In the pure crystalline phase monolayers, the neutralisation of electric dipoles by the formation of centrosymmetric dimers plays a significant role in determination of the monolayer configuration. This is also evident in the mixed monolayer structures where only homodimers in a centrosymmetric configuration can electrically

cancel the respective dipoles of the individual bases and domains of pure composition are formed (Figure 5). Electrical neutralisation must contribute significantly to the driving force for phase separation and in a completely energy minimised system, all molecules would be immobilised in domains of pure crystallinity. This would have little value as an information encoding mechanism. However, the aperiodic monolayer systems are not energy minimised and the structures can be trapped in local minima where aperiodic structure can exist, formed presumably from non-optimal interactions between bases as seen in Figure 5. Indeed, by supplying mechanical energy through the scanning action of the STM probe, the local aperiodic structure can be converted from one local minimum to another (Figures 5B, K).

The aperiodic structures could be viewed as imperfections from the pure phase-separated crystalline states and in that regard bear some resemblance to Cairns-Smith's inorganic 'crystal genes' (Cairns-Smith, 1982). The non-optimal interactions that comprise the aperiodic features are geometrically and energetically consistent with the observed crystalline configurations because the geometries are reasonably planar and the stabilisation energies are of the same order (Table I). However, they are not electrically neutralised. In these regions, electrical neutralisation, or minimisation might be achieved through complex arrangements of bases which could cancel out the dipoles of the individual bases in a convoluted way. A full interpretation of the aperiodic arrangement will require molecular modelling of clusters and higher level calculations of the *ab initio* type (Sponer *et al.*, 1996) which include electron correlation terms to accurately predict base-base geometries and to account for the electronic interactions of the molecules.

4.3. RELEVANCE OF THESE OBSERVATIONS TO PRIMITIVE CODING SYSTEMS

The monolayer structures shown here (Figure 5) are reminiscent of Schrödinger's description of the gene as an 'aperiodic solid' (Schrödinger, 1944) and of Cairns-Smith's inorganic 'crystal genes' (Cairns-Smith, 1982). The idea that inorganic interfaces could act as an informational reservoir through the presence of defects on the crystalline surface is attractive because of the aperiodicity of crystal defects and the presence of ionic sites within the crystal lattice which facilitate organic-inorganic electrostatic adsorption. The crystal gene proposal has received some experimental support from the observed polymerisation of nucleotides (Ferris and Ertem, 1992; Ferris *et al.*, 1996) and amino acids (Bujdak and Rode, 1996) in the presence of clays. The crystal gene model, however, has two obstacles. (1) Inorganic coding is complicated by the need for an additional evolutionary stage in which the inorganic template is replaced by an organic template and introduces the difficulty of preserving the coding properties during the transition. (2) Yockey (1992) has argued that such a mechanism of DNA-like information encoding in clay is mathematically impossible because 'the crystal patterns that Cairns-Smith shows exhibit far too much regularity'.

Our model removes these two obstacles. (1) The aperiodic structures are organic

and use, in the first instance, elements of the modern coding machinery that were potentially available in the prebiotic era. (2) The hydrogen bonding interactions between planar-arranged purine and pyrimidine bases and the number of interaction types will be discrete and determined by the hydrogen bond donor and acceptor geometries of the molecules involved. Combinatorial arrangements of the bases and the resultant structures of the aperiodic monolayers could provide, however, the physicochemical increase in complexity required for a primitive coding system.

Computer models of artificial life using cellular automata, which simulate 2-dimensional jigsaw-like systems dominated by lateral interactions between adjacent cellular components, have shown that information processing capabilities emerge spontaneously at the vicinity of a solid-fluid interface (Langton, 1986, 1989; Rasmussen *et al.*, 1991; Kauffman, 1993). This phase boundary constitutes a natural domain of information processing because the dynamics of the interface allows the storage, communication and transformation of information (Rasmussen *et al.*, 1991). Similarly, programmable matter (Toffoli and Magolus, 1991) provides a route to understanding the development of informational self-replication (Rasmussen *et al.*, 1991). Self-programming is the utilisation of local deterministic interactions between objects to generate information storage capability. Hydrogen bonding between the bases in the monolayers is just such a case because the base-base interactions are deterministically controlled by the geometry and physicochemistry of the bases. The STM images shown here (Figure 5) are direct evidence of aperiodic monolayer structure and, consequently, the self-programmable properties of purine and pyrimidine bases. Self-assembly of monolayers at the solid-fluid interface occurs within the domain of information processing and makes a connection between self-programming, information processing and monolayer formation tenable within the putative prebiotic environment.

However, the formation of aperiodic structure confers only the potential that information storage can occur. For the information of the monolayer patterns to have meaning, there must be some way to retrieve the information content (Yockey, 1992). We propose that the information content of the monolayers might be interpreted in terms of the coding for peptide catalysts. This hypothesis is consistent with cyano chemistry having yielded both bases and amino acids on the prebiotic earth (Ferris and Hagan, 1984).

For an informational mechanism to encode the primary order of amino acids in peptide synthesis, there must be a discrimination step in which the coding elements in the monolayers physically specify which amino acid is to be connected to the next. The adsorbed monolayer adds the functionality of a complex geometric and electronic corrugation potential to the surface because the purine and pyrimidine molecules are composed of chemically distinct mixtures of atoms which differ in their size and electronegativity. Some of these can act as proton acceptors and the molecular structures of the bases also contribute π -electrons from their aromatic arrangements and co-operative electronic interactions. In our model, the physicochemical differences in the structure of the surface corrugation acts as the chemical

discrimination mechanism because different combinations of bases would have different physicochemical properties.

The geometric placement of proton accepting groups in the monolayers are consistent with peptide bond dimensions and we have proposed that amino acids could directly interact, in a coded way, with the purine and pyrimidine bases embedded within the monolayers (Sowerby, 1995; Sowerby *et al.*, 1996; Sowerby and Heckl, 1998). In this model, the amino acids are zwitterions and the amino acid-monolayer interactions are mediated by hydrogen bonds formed between the amino acid ammonium groups with proton acceptors in the monolayer. The amino acids interact with the monolayer such that the ammonium and carboxylate groups of the amino acids constitute the interaction surface that is recognised by the monolayer. In this arrangement the ammonium and carboxylate groups are in-plane and in the correct orientation for peptide bond formation between juxtaposed molecules, the R-group points away from the surface and is not directly involved in the molecular discrimination.

Molecular recognition is dominated by molecular geometry and frontier orbital electron densities (Zheng *et al.*, 1995). The proposed discrimination of amino acids by the monolayer relies upon modulation of the frontier orbitals of the amino acids and the identification of these by the appropriate (amino acid-specific) molecular recognition motifs within the corrugation of the monolayer. The modulation is the result of the effect of the amino acid R-group on the molecular geometry and electron orbital electron densities of the ammonium and carboxylate groups of the amino acid and so different amino acids can be distinguished. The chemical nature of the amino acids has been investigated as a possible source of coding discrimination (Sjöström and Svante, 1985; Siemion and Stefanowicz, 1992) and supports amino acid discrimination at the ammonium and carboxylate functionalities that are not reliant on direct physical contact with the R-group. Chiral discrimination could also be achieved on enantiomorphic monolayer domains because the chiral centre of the amino acids is the bridging atom between the ammonium, carboxylate and R-groups and the proposed configuration of amino acid-monolayer interaction would allow for a simultaneous discrimination between amino acids based on R-group and stereochemistry.

Peptide bond formation is thermodynamically favoured at a solid-fluid interface (Orgel, 1998) and the putative polymerisation of amino acids on a monolayer template would give rise to encoded proteins. Although the information source in the aperiodic monolayer is two-dimensional, peptide products will still be one-dimensional since this is dictated by the amino acids which are joined together in a head-to-tail fashion due to the chemistry of their ammonium and carboxylate functional groups and the peptide bond they form. The solid-liquid interface facilitates peptide bond formation, presumably because a degree of rotational freedom is removed and this increases the chances of favourable ammonium-carboxylate collisions between amino acids. Any process that increases the local concentration of the amino acids will serve to catalyse this reaction. Additionally, because li-

quid water is sterically hindered at the interface, peptide bond condensation would be favoured over hydrolysis. Monolayer formation is also driven by purine and pyrimidine molecules seeking hydrogen bonding interactions at the interface where water is sterically hindered (De Levie and Wandlowski, 1994) and so, conveniently, both monolayer self-assembly and peptide bond synthesis will be driven by the, underlying, exclusion of water principles.

5. Summary

Contrived physical conditions were required to achieve the aperiodic structures that we have observed. Obtaining STM images of the aperiodic structures is not trivial. Often, the evaporation procedure results in phase separation with predominantly one monolayer phase observed. This is presumably related to the energetically favourable electrical neutralisation of the bases by the formation of centrosymmetric dimers and by differences in adsorption energies of the two components which, in the case of a purine and a pyrimidine, would be significantly different with the van der Waals component to adsorption being larger for the larger purine molecules. This is compounded by the preparation procedure, which concentrates the molecular species during the evaporation process and does not allow for well-characterised deposition parameters. The evaporation procedure was used here because it facilitates the removal of water from the interface to allow STM in air. In an origin of life scenario, however, it is also plausible that low concentrations of purine and pyrimidine bases would allow slow, salt-mediated monolayer growth processes at the solid-liquid interface and allow for aperiodic monolayer formation from a mixture of abiotic products. This has been most clearly demonstrated by real-time in situ STM of spontaneous adenine and guanine monolayer growth on graphite surfaces from aqueous saline solutions containing dilute concentrations of the bases (Tao and Shi, 1994b).

Despite experimental difficulties, it is clear that purine and pyrimidine bases have the potential to form aperiodic structures through spontaneous and putatively-prebiotic processes. From the point of view of the origin of life, the origin of the primordial genetic information could be inherently contained within the geometry and hydrogen bonding capabilities of the bases and, because the monolayers form spontaneously, a specific construction tool is not required. The genesis of primordial genetic information would come spontaneously from the self-programming capability of the purine and pyrimidine bases and monolayer structures that they would form. The monolayer-mediated construction of protein molecules capable of catalytic contributions to the synthesis of both purine and pyrimidine bases and amino acids would thus constitute a self-replicating chemical system with many of the hallmarks of modern biochemistry and outlines a framework for future investigations of both the physical system and with artificial life based on two-dimensional cellular automaton-like models.

Acknowledgements

We are grateful to Hubert P. Yockey and Michael J. Denton for informative discussions. This work was supported by the New Zealand Lottery Grants Board and a generous gift from Sir Roy McKenzie.

References

- Allen, M. J., Balooch, M., Subbiah, S., Tench, R. J., Balhorn, R. and Siekhaus, W.: 1992, *Ultramicroscopy* **42–44**, 1049.
- Allen, M. J., Balooch, M., Subbiah, S., Tench, R. J., Siekhaus, W. and Balhorn, R.: 1991, *Scanning Microsc.* **5**, 625.
- Anfinson, C. B.: 1973, *Science* **181**, 223.
- Barrett, S. D.: 1998, *Image SXM*, by Dr. S. D. Barrett, is a version of the public domain image analysis software NIH Image that has been extended to handle scanning microscope images from various manufacturers. It can be downloaded from <http://reg.ssci.liv.ac.uk/>.
- Baudisch, O.: 1913, *Zeitschrift der Angewandte Chemie* **26**, 612.
- Bernal, J. D.: 1951, *The Physical Basis of Life*, London, Routledge and Keegan Paul.
- Binnig, G., Rohrer, H., Gerber, C. and Weibel, E.: 1982, *Phys. Rev. Lett.* **49**, 57.
- Boland, T. and Ratner, B. D.: 1994, *Langmuir* **10**, 3845.
- Brack, A.: 1993, *Pure and Applied Chemistry* **65**, 1143.
- Bujdak, J. and Rode, B. M.: 1996, *J. Mol. Evol.* **43**, 326.
- Cairns-Smith, A. G.: 1982, *Genetic takeover and the mineral origins of life*, Cambridge University Press.
- Cairns-Smith, A. G.: 1988, *International Reviews in Physical Chemistry* **7**, 209.
- Dawson, R. M. C., Elliot, D. P., Elliot, W. H. and Jones, K. M.: 1986, *Data for Biochemical Research*, 3rd ed., Oxford, Oxford University Press.
- De Levie, R.: 1988, *Chem. Rev.* **88**, 599.
- De Levie, R. and Wandlowski, T.: 1994, *J. Electroanal. Chem.* **366**, 265.
- Edelwirth, M., Freund, J., Sowerby, S. J. and Heckl, W. M.: 1998, *Surf. Sci.* **417**, 201.
- Ferris, J. P. and Ertem, G.: 1992, *Science* **257**, 1387.
- Ferris, J. P. and Hagan, J. W. J.: 1984, *Tetrahedron* **40**, 1093.
- Ferris, J. P., Hill, A. R. J., Liu, R. and Orgel, L. E.: 1996, *Nature* **381**, 59.
- Freund, J., Edelwirth, M., Kröbel, P. and Heckl, W. M.: 1997, *Phys. Rev.* **B55**, 5394.
- Furukawa, M., Tanaka, H. and Kawai, T.: 1997, *Surf. Sci.* **417**, L33.
- Heckl, W. M.: 1993, *Rastertunnelmikroskopie an zweidimensionalen Kristallen aus organischen Molekülen*, Habilitation Thesis. Ludwig Maximilians Universität, Munich.
- Heckl, W. M., Smith, D. P. E., Binnig, G., Klagges, H., Hänsch, T. W. and Maddocks, J.: 1991, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8003.
- Hill, A. R. J., Böhler, C. and Orgel, L. E.: 1998, *Orig. Life Evol. Biosphere.* **28**, 235.
- Hoogsteen, K.: 1963, *Acta Crystallographa* **16**, 907.
- Jeffrey, G. A. and Saenger, W.: 1994, *Hydrogen bonding in biological structures*, Berlin, Springer-Verlag.
- Joyce, G. F.: 1989, *Nature* **338**, 217.
- Kauffman, S. A.: 1993, *The Origins of Order: Self-organisation and Selection in Evolution*, New York, Oxford University Press Inc.
- Kawai, T., Tanaka, H. and Nakagawa, T.: 1997, *Surf. Sci.* **386**, 124.
- Kistenmacher, T. J. and Shigematsu, T.: 1974, *Acta Crystallographa* **B30**, 1528.
- Langton, C. G.: 1986, *Physica* **22D**, 120.

- Langton, C. G.: 1989, *Artificial Life*, Langton, C. G. (ed.), Addison-Wesley.
- Löb, W.: 1913, *Berichte der deutschen chemischen Gesellschaft* **46**, 684.
- Mathews, F. S. and Rich, A.: 1964, *J. Mol. Biol.* **8**, 80.
- Miller, S. L.: 1953, *Science* **117**, 528.
- Nakagawa, T., Tanaka, H. and Kawai, T.: 1997, *Surf. Sci. Lett.* **370**, L1444.
- O'Brien, E. J.: 1967, *Acta Crystallographa* **23**, 92.
- Orgel, L. E.: 1998, *Origins of Life Evol. Biosphere.* **28**, 227.
- Park, S. and Quate, G. C.: 1986, *Appl. Phys. Lett.* **48**, 112.
- Parry, G. S.: 1953, *Acta Crystallographa* **7**, 313.
- Poler, J. C., Zimmerman, R. M. and Cox, E. C.: 1995, *Langmuir* **11**, 2689.
- Rasmussen, S., Knudsen, C. and Fieldberg, R.: 1991 *Dynamics of Programmable Matter. Artificial Life II, SFI Studies in the Sciences of Complexity*, Langton, C. G. et al. (eds.), Redwood City, CA, Addison-Wesley, X, p. 211.
- Reggia, J. A., Armentrout, S. L., Chou, H. and Peng, Y.: 1993, *Science* **259**, 1282.
- Sakore, T. D. and Sobell, H. M.: 1969, *J. Mol. Biol.* **43**, 77.
- Schrödinger, E.: 1944, *What is Life?*, Cambridge University Press.
- Schwartz, A. W. and de Graaf, R. M.: 1993, *J. Mol. Evol.* **36**, 101.
- Siemion, I. Z. and Stefanowicz, P.: 1992, *Biosystems* **27**, 77.
- Sjöström, M. and Svante, W.: 1985, *J. Mol. Evol.* **22**, 272.
- Smith, J. V.: 1998, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3370.
- Sowerby, S. J.: 1995, Ph.D. Thesis. Otago, New Zealand.
- Sowerby, S. J., Edelwirth, M. and Heckl, W. M.: 1998a, *Appl. Phys. A* **66**, S649.
- Sowerby, S. J., Edelwirth, M. and Heckl, W. M.: 1998b, *J. Phys. Chem. B* **102**, 5914.
- Sowerby, S. J., Edelwirth, M., Reiter, M. and Heckl, W. M.: 1998c, *Langmuir* **14**, 5195.
- Sowerby, S. J. and Heckl, W. M.: 1998, *Orig. Life Evol. Biosphere.* **28**, 283.
- Sowerby, S. J., Heckl, W. M. and Petersen, G. B.: 1996, *J. Mol. Evol.* **43**, 419.
- Sowerby, S. J. and Petersen, G. B.: 1997, *J. Electroanal. Chem.* **433**, 85.
- Sowerby, S. J. and Petersen, G. B.: 1999, *Orig. Life Evol. Biosphere* (in press).
- Sponer, J., Leszczynski, J. and Hobza, P.: 1996, *Journal of Biomolecular Structure and Dynamics* **14**, 117.
- Srinivasan, R. and Gopalan, P.: 1993, *J. Phys. Chem.* **97**, 8770.
- Srinivasan, R. and Murphy, J. C.: 1992, *Ultramicroscopy* **42-44**, 453.
- Srinivasan, R., Murphy, J. C., Fainchtein, R. and Pattibiraman, N.: 1991, *J. Electroanal. Chem.* **312**, 293.
- Stupian, G. W. and Leung, M. S.: 1987, *Appl. Phys. Lett.* **51**, 1560.
- Tanaka, H. and Kawai, T.: 1995, *Materials Science and Engineering C* **3**, 143.
- Tanaka, H., Nakagawa, T. and Kawai, T.: 1996a, *Surf. Sci.* **364**, L575.
- Tanaka, H., Yoshinobu, J., Kawai, M. and Kawai, T.: 1996b, *Japanese Journal of Applied Physics* **35**, L244.
- Tao, N. J., DeRose, J. A. and Lindsay, S. M.: 1993, *J. Phys. Chem.* **97**, 910.
- Tao, N. J. and Shi, Z.: 1994a, *J. Phys. Chem.* **98**, 7422.
- Tao, N. J. and Shi, Z.: 1994b, *J. Phys. Chem.* **98**, 1464.
- Tao, N. J. and Shi, Z.: 1994c, *Surf. Sci. Lett.* **301**, L217.
- Tao, N. J. and Shi, Z.: 1994d, *Surf. Sci. Lett.* **321**, L149.
- Toffoli, T. and Magolus, N.: 1991, *Physica D* **47**, 263.
- Vetterl, V.: 1965, *Experientia* **21**, 9.
- Vetterl, V.: 1966, *Collect. Czech. Chem. Commun.* **31**, 2105.
- Voet, D. and Rich, A.: 1969, *J. Am. Chem. Soc.* **91**, 3069.
- Weimer, M., Kramar, C. B. and Baldeschwieler, J. D.: 1988, *Phys. Rev. B* **37**, 4292.
- Yockey, H. P.: 1992, *Information Theory and Molecular Biology*, Cambridge University Press.
- Zheng, Y., Neville, J. J. and Brion, C. E.: 1995, *Science* **270**, 786.