Orig Life Evol Biosph (2008) 38:399–418
DOI 10.1007/s11084-008-9143-4

THEORETICAL MODELING

Compositional Inheritance: Comparison of Self-assembly and Catalysis

Meng Wu · Paul G. Higgs

Received: 21 February 2008 / Accepted: 26 June 2008 /
Published online: 18 July 2008

Abstract Genetic inheritance in modern cells is due to template-directed replication of nucleic acids. However, the difficulty of prebiotic synthesis of long information-carrying polymers like RNA raises the question of whether some other form of heredity is possible without polymers. As an alternative, the lipid world theory has been proposed, which considers non-covalent assemblies of lipids, such as micelles and vesicles. Assemblies store information in the form of a non-random molecular composition, and this information is passed on when the assemblies divide, i.e. the assemblies show compositional inheritance. Here, we vary several important assumptions of previous lipid world models and show that compositional inheritance is relevant more generally than the context in which it was originally proposed. Our models assume that interaction occurs between nearest neighbour molecules only, and account for spatial segregation of molecules of different types within the assembly. We also draw a distinction between a self-assembly model, in which the composition is determined by mutually favourable interaction energies between the molecules, and a catalytic model, in which the composition is determined by mutually favourable catalysis. We show that compositional inheritance occurs in both models, although the self-assembly case seems more relevant if the molecules are simple lipids. In the case where the assemblies are composed of just two types of molecules, there is a strong analogy with the classic two-allele Moran model from population genetics. This highlights the parallel between compositional inheritance and genetic inheritance.

Keywords Autocatalytic set · Catalysis · Compositional inheritance · Lipid vesicle · Lipid world · Moran model · Self-assembly

Introduction

All cellular life has three essential components (Ganti 1979; Szathmary et al. 2005). A genetic system is required to store information and to pass that information on in a heritable

M. Wu · P. G. Higgs (✉)
Origins Institute and Department of Physics and Astronomy, McMaster University, Hamilton, ON L8S 4M1, Canada
e-mail: higgsp@mcmaster.ca

way. A metabolic system is required to synthesize the necessary molecules and generate energy in a form that is usable by the organism. A boundary system (such as a cell membrane) is required to distinguish individuals from each other and the environment. Only when we have distinct individuals can multiplication, competition and evolution occur. However, the simplest conceivable cell possessing all three components is already rather complex; therefore when considering the origin of life, it is natural to ask whether something simpler could have existed that did not possess all three.

One approach to the origin of life focuses on the origin of informational polymers. DNA is the genetic polymer in modern cells and proteins are the catalysts that regulate the majority of metabolic reactions. However, it is likely that there was an RNA world period in the early history of life on earth in which RNA was the informational polymer and also the key catalytic molecule (Gilbert 1986; Jeffares et al. 1998; Joyce 2002; Orgel 2004). The main evidence for this is that the sequences of proteins in modern cells are encoded in mRNAs, that rRNAs are the main functional component of the ribosome, and tRNAs enable translation to occur via the codon–anticodon interactions that define the genetic code. Several other types of RNAs that may be relics of the RNA world are also known. We therefore envisage an RNA world in which there were RNA-based cells containing a fairly diverse set of functional RNA molecules. The origin of translation and the genetic code would have occurred in such RNA-based cells (Higgs and Pudritz 2007).

Cells in a late-stage RNA world would have possessed all three of the essential components. What preceded this stage? A stronger version of the RNA world hypothesis is that life originated in the form of self-replicating RNA molecules. Considerable progress has been made in the in vitro synthesis of polymerase ribozymes that are able to synthesize oligomers in a template-directed manner (Bartel and Unrau 1999; Johnston et al. 2001; Zaher and Unrau 2007). However, it is still an open question whether it will eventually be possible to synthesize ribozymes capable of self-replication. If such ribozymes did exist at the time of the origin of life, they need not necessarily have been contained in cells. Szabo et al. (2002) have studied a model of a population of sequences that cooperate with one another to make duplicate copies of the neighbouring sequences. However, naked ribozyme sequences like this would have to exist in an environment in which single nucleotides were readily available and in which polymerization of RNA was thermodynamically favourable. Thus although there seems to be no problem with this picture from the point of view of evolutionary theory, it may be that it is ruled out by chemistry and thermodynamics. Shapiro (1988, 2006) has argued against the idea of self-replicating ribozymes due to the difficulty of synthesis and the lack of stability of RNA in prebiotic conditions. Orgel (2004) also concludes that we are far from having a plausible prebiotic synthesis of all the components of nucleotides and a way of reliably synthesizing long RNA polymers.

If the formation of information-carrying polymers is difficult as an initial step, then is it possible to have life that would have a metabolic system and/or a membrane system but no genetic system? Several authors have proposed ‘metabolism first’ theories in which a self-sustaining network of chemical reactions arises in absence of informational polymers or cells. It is presumed that a ‘food’ set of small molecules is present in the environment and is maintained at a high concentration. These molecules may react together in many ways to produce a huge variety of larger molecules. Spontaneous reactions are assumed to be slow, but some of the larger molecules have the ability to catalyze some of the possible reactions and make them proceed much faster than the spontaneous rate. An autocatalytic set of the larger molecules may arise, i.e. each molecule in the autocatalytic set can be formed by reactions involving the food set and other members of the autocatalytic set, and each formation reaction is catalyzed by another molecule in the set. Kauffman (1986) considered

a model in which the possible molecules are copolymers made from a small number of monomeric units, and the possible reactions are ligation and hydrolysis of these copolymers. The food set is the set of oligomers of length shorter than some initial length L. Autocatalytic subsets of long polymers are found to arise frequently in the model if the food set is sufficiently diverse. This model was initially formulated in terms of the reaction graph only, but was later extended to include rates and concentrations (Farmer et al. 1986; Bagley and Farmer 1991). Although Kauffman (1986) refers to the polymers as proteins, the model is defined in an abstract way and could just as well apply to autocatalytic sets of RNAs. However, template-directed synthesis is not considered, and the reactions apply to specific sequences rather than to any sequence (as with a universal polymerase ribozyme). Thus, these theories represent a metabolism-first view not a replication-first view, whatever the nature of the chemical reactants is thought to be.

Another example of an autocatalytic set model is the graded autocatalysis replication domain (GARD) model of Segré et al. (1998). The food set consists of a large number of possible monomers. Dimers can form between any two types of monomers. Some of the dimerization reactions are catalyzed by some of the dimers. Hence, an autocatalytic set of dimers may emerge. The lipid world model (Segré et al. 2000, 2001a, b) was originally developed as a simplification of the GARD model. In the lipid world model there is a diverse set of molecules available in the environment. These molecules do not react chemically with one another, but they can form non-covalent assemblies, such as micelles and vesicles, due to the physical interactions between them. New molecules can be added to assemblies and existing molecules can be removed. The rate of addition and removal are dependent on the other molecules in the assembly. The assemblies contain subsets of molecules that together promote the addition of more molecules of the same types – in this sense they are an autocatalytic set. Random splitting of the assemblies occurs when they reach a specified maximum size. Thus, the model has growth and division of entities with a boundary system but no metabolism or genetic system.

The most important aspect of the lipid world model is that it demonstrates compositional inheritance (Segré et al. 2000, 2001a). Even though the assemblies do not possess informational polymers, they do possess information in the form of the non-random composition of the molecules of which they are comprised. If this compositional information is stable over time and is passed on to the two descendant assemblies when splitting occurs, then the system possesses compositional inheritance. The lipid world model highlights this point by removing all the complexities of the chemical reactions. We note that the earlier model of Dyson (1999) considers a molecular assembly that can exchange molecules with the external environment. This system maintains a non-random molecular composition, but growth and division are excluded, so in our view, there is no inheritance in Dyson's model. The autocatalytic copolymer models (Kauffman 1986; Farmer et al. 1986) and the GARD model (Segré et al. 1998) also do not directly demonstrate compositional inheritance because they consider only one reaction network rather than a population. However, if metabolism-first theories for the origin of life are correct, then the metabolic reactions have to become associated with compartmentalized individuals that grow and divide, giving rise to populations of protocells containing autocatalytic reactions sets but no informational polymers. Such a population would also have to have compositional inheritance so that the same autocatalytic metabolism would be passed on.

Thus, we think that the idea of compositional inheritance is an important concept that goes beyond the original definition of the lipid world model. In this paper we will consider models for the growth and division of populations of molecular assemblies in which the

assumptions of the lipid world model are changed in several ways. If the phenomenon of compositional inheritance can be shown to be robust to these changes, this will strengthen the argument that some form of compositional inheritance was an important ingredient in the origin of life.

Catalysis versus Self-assembly

Before defining the specific models that we will study here, we wish to emphasize that the word 'catalysis' is used in several different ways in the studies mentioned above. In the usual sense of the word, a catalyst is a molecule that speeds up a chemical reaction without taking part in the reaction and without being used up in the process. The molecules in the autocatalytic copolymer models and the GARD model are catalysts in this sense. In the model of Bagley and Farmer (1991), all polymers of a given length are equivalent thermodynamically and they would all have the same concentration if there were no catalysis or if the system were allowed to proceed to equilibrium. However, when an autocatalytic set arises, the formation of the molecules in the set is catalyzed, and the concentrations of these sequences reach a much higher value than those of the sequences not in the set. This is called 'catalytic focusing'. The concentrations remain out of equilibrium because of continual input and exit of molecules from the system. The choice of polymers that are included in the autocatalytic set is determined by the catalytic properties of the molecules, not thermodynamic properties.

There are also cases where lipids can catalyze chemical reactions. Bachmann et al. (1992) studied an autocatalytic system in which micelles composed of sodium caprylate can catalyze the hydrolysis of ethyl caprylate to form more sodium caprylate, which forms more micelles. Fellerman and Solé (2007) also carried out a simulation of lipid aggregation coupled to catalyzed lipid formation. Walde (2006) gives many examples of chemical reactions that are catalyzed by the presence of lipid micelles and vesicles. However, in the context of the lipid world, the word catalysis is often used in a different way. Szathmary et al. (2005) state that 'it is important to point out that membrane growth is an autocatalytic process.' This is correct in the sense that the presence of the membrane promotes the addition of further molecules to the membrane, but in this case it is thermodynamics that drives the addition of the new molecules, not catalysis. The reason lipids form micelles and membranes is because of the interactions between them – the hydrophobic tails have a lower energy of interaction with one another than with water, while the hydrophilic head groups have a lower energy of interaction with the water. Moving a single molecule of lipid from solution to the membrane lowers the free energy of the system. We feel that 'self-assembly' is a better word for what is happening when lipids form micelles or membranes, and that 'catalysis' should be avoided in this context.

The term catalysis is also used in cases where there is a barrier between two states, and the catalyst acts to lower this barrier. An example like this in modern cells is a membrane protein that forms a channel though which a charged molecule can cross a membrane. The membrane would otherwise be almost impermeable to the charged molecule. The membrane protein clearly speeds up the configurational change without being used up in the process, and it may be legitimate to call it a catalyst. However, molecules as complex as membrane proteins were not around at the time envisaged in the lipid world model. In the lipid world models of Segré et al. (2000, 2001a, b), catalysis is used to refer to speeding up the addition or removal of a molecule. It seems likely that a new lipid would be added rapidly because it is attracted to the other molecules in the assembly, rather than because

there is a barrier for addition that is somehow lowered by the other molecules. Therefore the 'barrier-crossing' usage of catalysis also does not seem very appropriate for lipid assemblies.

This distinction is not just a matter of words. It becomes important when we consider the rates of addition and removal of molecules of different types. If a molecule of a certain type has a favourable energetic interaction with the existing molecules in the assembly then its rate of addition should be rapid. Once it is added to the assembly, its rate of removal should be slow because there will be an increase in energy going in the reverse direction. Conversely, if a molecule of another type has an unfavourable energetic interaction with the existing molecules then its rate of addition will be slow and its rate of removal will be fast, if it is added. However, this is *not* what happens in the kinetic equations used in previous lipid world models (equation 4 of Segré et al. 2000, and equation 10 of Segré et al. 2001a). According to these equations, if the rate of addition of one type of molecule is fast, then the removal is also fast, or alternatively, if the addition rate of another type of molecule is slow, the removal rate is slow. These equations represent catalysis in the sense that both forward and backward reactions are speeded up equally. As a result, the composition of the assembly is determined by catalytic focusing, as in the copolymer autocatalytic set models. This would make sense if we were describing something like the membrane protein above, because the protein would speed up the passage of the charged molecule in both directions, but it seems less reasonable to us in the case of formation of lipid micelles and membranes. We expect the differences in the thermodynamic properties of the molecules to be more important in determining the constituents of an assembly than the differences in their catalytic properties. Therefore, in this paper we define a model of self-assembly in which the rates of additional and removal of molecules are controlled by their thermodynamic properties.

Another detail of the equations of Segré et al. is that the rate of addition of a molecule depends on the concentration of all the other molecules in the assembly. This means that the new molecule mixes freely with all the other molecules and is able to interact with all of them. It is more reasonable to suppose that a molecule will only interact with its spatial neighbours. The spatial arrangement of molecules is likely to be important because the differences in thermodynamic interactions between the molecules will lead to attraction or repulsion of molecules of different types. We will therefore include the effects of spatial arrangement and segregation of molecules in the models we discuss here.

Model Definitions

We suppose that K different types of molecules are present in the external environment from which assemblies may form. The total concentration of molecules is C, and the concentration of each type is C/K. We consider a population of N individual assemblies. Assemblies may grow or shrink by exchange of molecules with the environment. When the number of molecules in an assembly reaches a maximum size M, the assembly splits into two offspring of size $M/2$. At this point, another randomly chosen individual is removed from the population so that N stays fixed. The molecules in each assembly are arranged in a ring, i.e. a one-dimensional array with periodic boundaries. Each molecule interacts only with its two neighbours in the ring. This is the simplest structure in which the spatial arrangement of the molecules is relevant. It is intended as a contrast to the model of Segré et al. (2000) in which all molecules in an array interact with one another simultaneously. When a ring splits, one break point is chosen at random and the other break point is chosen

to be on the opposite side of the ring. The two half-size rings are then reconnected by establishing periodic boundaries within each one.

Let i, j, k and l label types of molecules chosen from the K possible types. Let $r_{\mathrm{add}}\,(k \mid ij)$ be the rate of addition of a type k molecule between molecules of types i and j (Fig. 1a), and let $r_{\mathrm{rem}}\,(k \mid ij)$ be the rate of removal of a type k molecule from between molecules of types i and j (Fig. 1b). We also allow diffusion of molecules within an assembly by exchange of neighbouring molecules. Let $r_{\mathrm{dif}}\,(jk \mid il)$ be the rate of exchange of molecules of types j and k between molecules of types i and l (Fig. 1c). We now define two models: a self-assembly model, in which the rates depend on interaction energies, and a catalytic model, in which the rates depend on catalytic properties.

Self-assembly model In this case there is a matrix ε_{ij} of energies of interaction between pairs of molecules of different types. The energy of an assembly is the sum of the interaction energies of all pairs of neighbours in the ring. The change in energy when the molecule of type k is added in Fig. 1a is $\Delta E_{\mathrm{add}} = \varepsilon_{ik} + \varepsilon_{kj} - \varepsilon_{ij}$. The change in energy when the molecule of type k is removed in Fig. 1b is $\Delta E_{\mathrm{rem}} = -\Delta E_{\mathrm{add}}$. The change in energy when molecules of types j and k exchange positions in Fig. 1c is $\Delta E_{\mathrm{dif}} = \varepsilon_{ik} + \varepsilon_{jl} - \varepsilon_{ij} - \varepsilon_{kl}$. To satisfy detailed balance, we require that the ratio of addition and removal rates be

$$\frac{r_{\mathrm{add}}(k|ij)}{r_{\mathrm{rem}}(k|ij)} = \frac{C}{K}\exp(-\Delta E_{\mathrm{add}}/k_{\mathrm{B}}T), \tag{1}$$

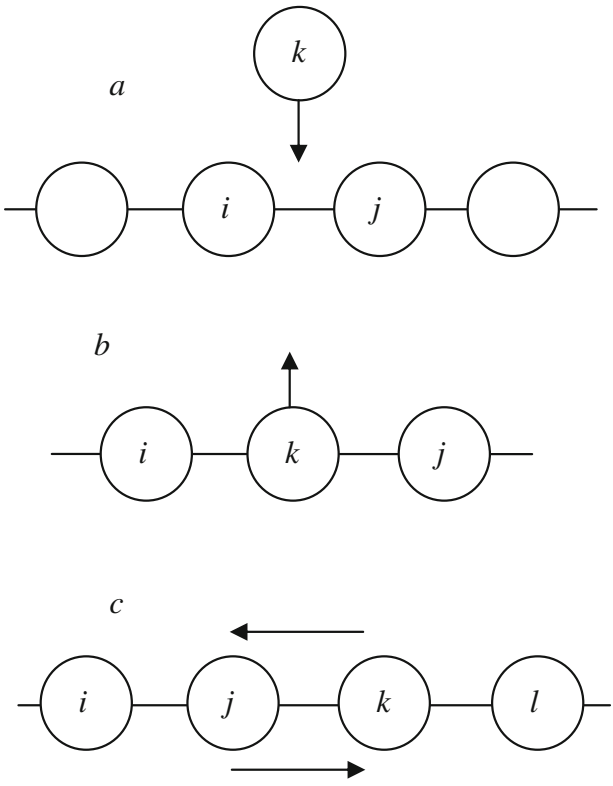

Fig. 1 **a** Addition of a type k molecule between molecules of types i and j. **b** Removal of a type k molecule from between molecules of types i and j. **c** Diffusive exchange of position of molecules of types j and k between molecules of types i and l

where k_B is Boltzmann's constant and T is the absolute temperature. Although this ratio is defined by thermodynamics, there is still flexibility in how the absolute values of the rates are defined. It is standard in Monte Carlo simulations of physical systems to set 'downhill' reactions to rate 1, and the corresponding uphill reactions to be slower by a factor of $\exp(-\Delta E/k_B T)$. However, if we do this, all reactions that are energetically favourable will occur at the same rate. Instead, we will define the rates in the following way:

$$\begin{aligned} r_{\text{add}}(k|ij) &= \frac{C}{K}\exp(-\Delta E_{\text{add}}/2k_B T), \\ r_{\text{rem}}(k|ij) &= \exp(+\Delta E_{\text{add}}/2k_B T) = \exp(-\Delta E_{\text{rem}}/2k_B T). \end{aligned} \tag{2}$$

With these choices, the reactions that are most energetically favourable will occur fastest. The more strongly attracted a molecule is to the assembly, the faster it will add to the assembly and the slower it will be removed. Often reaction kinetics is described in terms of transition states – in this case, a half-inserted molecule. The rates above assume either that there is no barrier associated with the half-inserted molecule, or that the barrier is systematically lower for molecules that have a lower free energy when fully inserted.

The ratio of the rate of exchange of j and k to the reverse exchange must also satisfy detailed balance:

$$\frac{r_{\text{dif}}(jk|il)}{r_{\text{dif}}(kj|il)} = \exp(-\Delta E_{\text{dif}}/k_B T). \tag{3}$$

We define these rates as:

$$r_{\text{dif}}(jk|il) = D\exp(-\Delta E_{\text{dif}}/2k_B T), \quad r_{\text{dif}}(kj|il) = D\exp(+\Delta E_{\text{dif}}/2k_B T), \tag{4}$$

where D is a constant that controls the relative rate of diffusion to addition and removal.

Catalytic Model In this case, the rates depend on a matrix β_{ij} of catalytic effects instead of a matrix of interaction energies:

$$r_{\text{add}}(k|ij) = \frac{C}{K}\left(1+\beta_{ik}+\beta_{jk}\right), \quad r_{\text{rem}}(k|ij) = \left(1+\beta_{ik}+\beta_{jk}\right). \tag{5}$$

This is similar to the model of Segré et al. because the rates of addition and removal of molecule k are affected by catalysis in the same way, so the ratio of addition and removal rates is independent of the type of molecule. However, we keep track of spatial position. The molecule is influenced by its two neighbours only, rather than by the average of the β factors of all the molecules in the assembly, as was the case for Segré et al. (2000, 2001a). We suppose that the diffusion is not affected by catalysis. Therefore, we set $r_{\text{dif}}(jk|il) = D$, independent of the type of molecules exchanged.

Implementation When simulating these models, the processes that occur need to be chosen stochastically with rates given by the above rules. The implementation of the stochastic simulation described here applies for either of the two models above, although the rates are defined in different ways. By 'site,' we refer to the position of one molecule in an assembly. The number of sites in an assembly is the number of molecules that it currently contains. At each Monte Carlo step, one site is chosen at random from the set of all the sites in all the

assemblies in the population. Let R_{site} be the sum of the rates of all the processes that could occur at a site. This can be written as

$$R_{\text{site}} = r_{\text{rem-site}} + \sum_k r_{\text{add-site}}(k) + r_{\text{dif-site}}, \tag{6}$$

where $r_{\text{rem-site}}$ is the rate of removal of the molecule currently at that site, $r_{\text{add-site}}(k)$ is the rate of addition of a molecule of type k between the molecule at this site and its neighbour to the right, and $r_{\text{dif-site}}$ is the rate of exchange of the molecule at this site with its neighbour to the right. These rates depend on the types of molecules at the site and the neighbouring sites according to the rules given above for either of the two models. Let $R_{\max}$ be the maximum possible value of R_{site}. As there is a finite number K of possible types of molecule, there is a finite number of local configurations, and $R_{\max}$ can be determined straighforwardly. Having randomly chosen the site, one process may occur at this site. With probability $r_{\text{rem-site}}/R_{\max}$ the molecule is removed; with probability $r_{\text{add-site}}(k)/R_{\max}$ a molecule of type k is added; and with probability $r_{\text{dif-site}}/R_{\max}$ two neighbouring molecules are exchanged. In the case where $R_{\text{site}} < R_{\max}$ there is a probability $1-R_{\text{site}}/R_{\max}$ that nothing happens at this Monte Carlo step.

Results – Self-assembly Model

We consider the simplest case with only two kinds of lipids, labelled 1 and 2, and define the interaction energy in terms of an energy parameter a: $\varepsilon_{11} = \varepsilon_{22} = -a$ for molecules of the same type, and $\varepsilon_{12} = \varepsilon_{21} = +a$ for molecules of different types. Note that if some non-zero mean interaction energy were added to all the ε_{ij}, this would not change the model because this factor could be incorporated into C in Eqs. 1 and 2. Therefore we will keep the mean ε_{ij} to be zero. The rates are therefore a function of a single energy ratio $\sigma = a/k_{\text{B}}T$.

The following simulations were performed with populations of $N=100$ individual assemblies, using a maximum size of $M=40$ molecules per assembly. We chose parameters $C=1.0$ and $D=1.0$. The simulations began with all individuals containing 20 molecules, each chosen to be 1 or 2 with equal probability. The simulation was run for an initial period to reach a stationary state. After this period, we recorded information on the composition of each assembly at the point of division. $P(z)$ is the probability that an individual has a fraction z of type 1 molecules when it reaches its maximum size M prior to division. Although z is always a multiple of $1/M$, in Fig. 2, z is treated as a continuous variable, and the curve $P(z)$ is treated as a continuous probability distribution normalized so that there is unit area under the curve. If σ is small, most individuals have a mixed composition (z close to 0.5). If σ is large, most individuals are dominated by either type 1 or type 2 molecules (z close to 0 or 1).

The function $P(z)$ measures variability of molecular composition of individuals. We also want to measure variability at the population level. To do this, we classify individuals as either type 1 or type 2, according to whether they contain a majority of type 1 or type 2 molecules. In the rare case that the two types have equal numbers, the individual is classed as type 1 or type 2 randomly. We then define x as the frequency of type 1 individuals in the population at any given instant in time. The function $\Phi(x)$ is the probability distribution of x. This is obtained by measuring x at regular intervals in the simulation and plotting a histogram of the stored values. The distribution is normalized as though x were a continuous variable, so that there is unit area under the $\Phi(x)$ curve. In Fig. 3, it can be seen that if σ is small, most populations have a mixed composition (x close to 0.5). If σ is large, most populations are dominated by either type 1 or type 2 individuals (x close to 0 or 1).

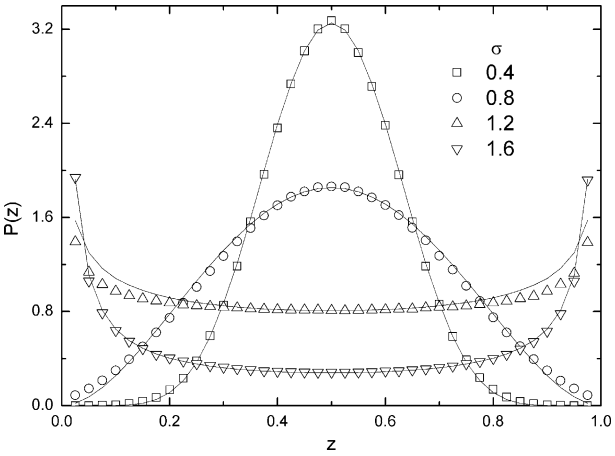

Fig. 2 Self-assembly model. Distribution $P(z)$ of the fraction of type 1 molecules in an individual at the point where it reaches maximum size. *Symbols* show simulation results with various values of the energy parameter σ. *Curves* are fits of the beta distribution to the data

The bell-shaped and U-shaped distributions in Figs. 2 and 3 will be familiar to population geneticists. They are examples of the beta distribution:

$$\Phi(x) = \frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)} x^{\theta-1}(1-x)^{\theta-1}. \tag{7}$$

The Moran model in population genetics considers the frequency distribution of alternative alleles at a given locus in a population in which genetic drift occurs due to birth and death events (Moran 1958; Karlin and McGregor 1962; Ewens 2004). If there are two alleles with an equal rate of mutation u in both directions between them, then the distribution of frequencies of either allele is given by equation 7. Allele frequency distributions are either bell-shaped or U-shaped, depending on the value of θ, which is $2Nu$ in the Moran model. In Fig. 3, we fitted Eq. 7 to the $\Phi(x)$ curve from the simulation of the self-assembly model, using θ as a single fitting parameter for each curve. This function fits the data very well. In our case, θ depends on σ rather than u in the Moran model. Large σ corresponds to small u, where the distributions are U-shaped, and lower σ corresponds to

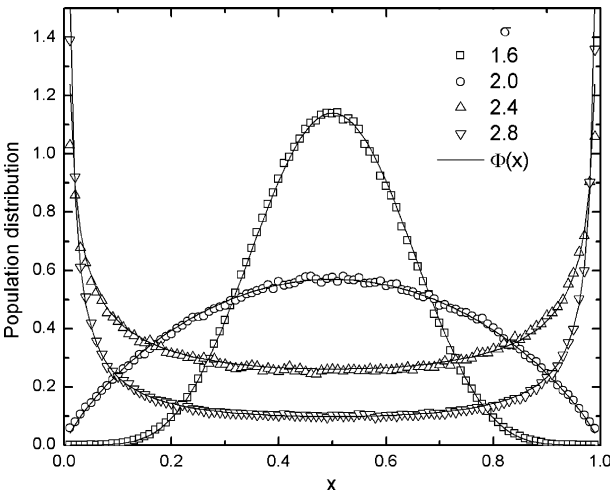

Fig. 3 Self-assembly model. Distribution $\Phi(x)$ of the fraction of type 1 individuals in the population. *Symbols* show simulation results with various values of the energy parameter σ. *Curves* are fits of the beta distribution to the data

higher u, where the distributions are bell-shaped. In Fig. 2, we used the same beta distribution to fit the $P(z)$ distribution. The function also fits the data quite well. $P(z)$ becomes U-shaped at smaller values of σ than $\Phi(x)$. For example, with σ=1.6, $P(z)$ is U-shaped but $\Phi(x)$ is still bell-shaped. Thus, for any value of σ, two different θ values are required to fit the two curves. The fitted values of θ are shown as a function of σ in Fig. 4.

To better understand the analogy between our model and the Moran model, it is necessary to show that there is heredity in our model. We therefore consider the relationship between compositions of parents and offspring. Let z_{parent} be the composition of a parent individual at the point when it reaches size M and is about to divide. Let $z_{\text{offspring}}$ be the composition of one of the offspring of this parent at some time later when it has also grown to size M and is about to divide. In the simulations, pairs of parent and offspring compositions were recorded over a long period of time. The contour plots in Fig. 5 show the probability distribution $p(z_{\text{offspring}} \mid z_{\text{parent}})$. Horizontal cross sections through the plots are normalized to 1, so that each cross section is a probability distribution for a given z_{parent}.

For σ=0.4, there is little correlation between parent and offspring. All offspring have composition close to z=0.5, irrespective of the parent, which means a random mixture of the two types of molecule. In contrast, for σ=0.8 and 1.2, there is a clear correlation. The mean value of $z_{\text{offspring}}$ increases as z_{parent} increases and is fairly close to z_{parent}. There is compositional inheritance in these cases; however, the distribution of offspring compositions is fairly broad, so the heredity is weak. For σ=1.6, heredity is strong. As $P(z)$ is U-shaped, most parents have compositions close to 0 or 1. The offspring also usually have composition close to 0 or 1, i.e. almost all the weight of the contour plot is in the two corners. The 'spottiness' of the top and bottom of the σ=0.4 plot close to z_{parent}=0 and 1 is statistical noise arising from the way the probability is normalized. As $P(z)$ is bell-shaped for this value of σ (see Fig. 2), there are very few parent individuals with these extreme compositions, so there is very little data contributing to these parts of the plot.

We define the following correlation function I as a quantitative measure of compositional inheritance:

$$I = \frac{\langle z_{\text{offspring}} z_{\text{parent}} \rangle - \langle z \rangle^2}{\langle z^2 \rangle - \langle z \rangle^2}. \tag{8}$$

I=1 if offspring and parents always have identical composition, and I=0 if offspring and parent compositions are uncorrelated. Figure 4 shows that I increases with σ, and is close to

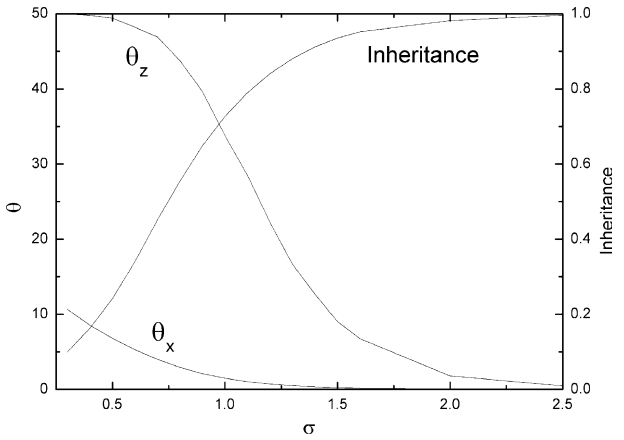

Fig. 4 Self-assembly model. *Left hand scale* shows the fitted values of θ for $P(z)$ and $\Phi(x)$ (denoted θ_z and θ_x) as a function of σ. *Right hand scale* shows the inheritance function I

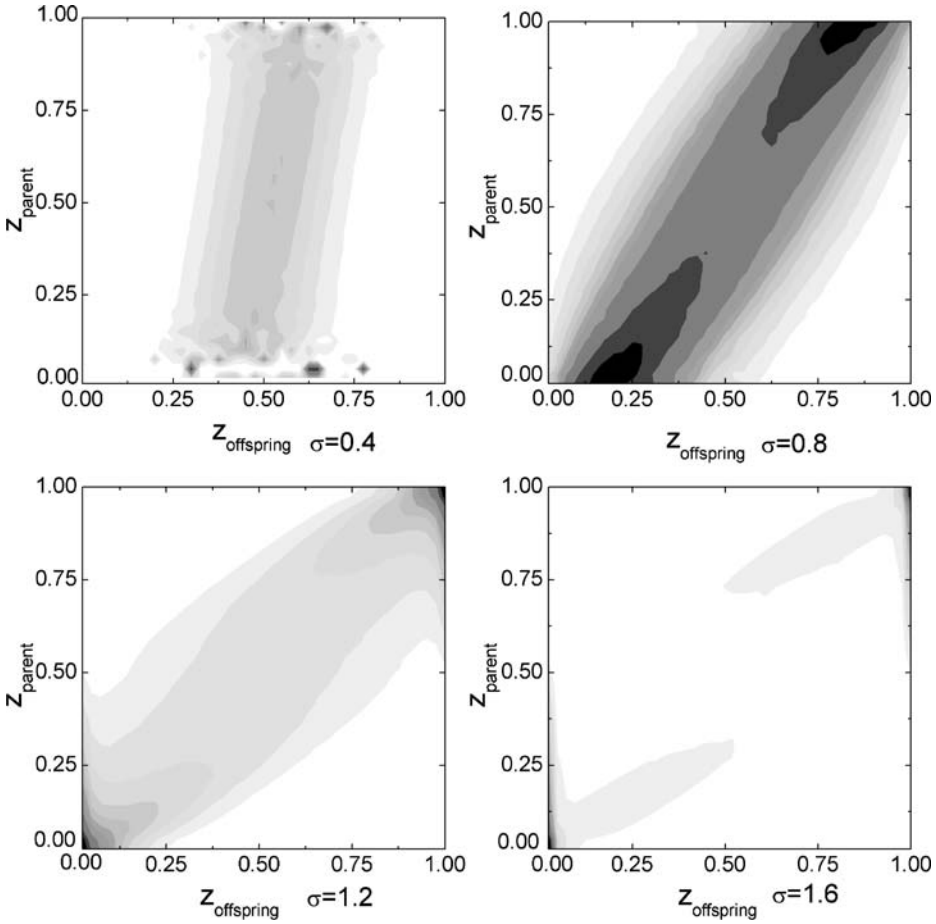

Fig. 5 Self-assembly model. Probability distributions of the offspring composition $z_{\text{offspring}}$ conditional on the composition of their parent z_{parent}

1 for σ larger than about 1.5. In this regime, individuals are usually of one extreme composition or other, and offspring are almost always of the same type as their parent.

The analogy between the self-assembly model with two types of molecules and the Moran model with two alleles is quite a strong one. At each time step in the Moran model a single individual multiplies and a single individual dies. In our model the birth and death events are controlled by the growth of the assemblies and do not occur regularly every time step. Nevertheless, when one individual splits, another individual is removed from the population so that the number of individuals remains constant, as in the Moran model. Therefore our model is also a birth and death process. Usually an assembly that is dominated by one type of molecule will grow and divide to produce two offspring dominated by the same type of molecule. Occasionally, a domain of the other type of molecule may be nucleated and grow, and the assembly may end up being dominated by the other type of molecule. This is the equivalent of a mutation from one type of assembly to another. For this change to occur, the assembly has to pass through high energy configurations with mixed composition. The energy barrier is higher when σ is higher. Hence, the effective mutation rate is lower, and the values of θ needed to fit the $P(z)$ and $\Phi(x)$ distributions are lower.

With the energies defined as above, molecules of the same type attract one another. Thus within each assembly, the molecules will not be randomly ordered. Instead, there will be alternating domains of molecules of the two types. The larger σ, the larger the typical size of these domains will become. For large enough σ, the typical domain size will exceed the finite size of the assembly itself. Thus, the whole assembly will become a single domain of one type of molecule. The energy rules in this model correspond to the one dimensional Ising model or binary alloy model (Yeomans 1992). In Fig. 6, we considered a 1d Ising model on a ring of 40 sites with the same energy rules as used to define the rates. We calculated the partition function exactly numerically, and determined the probability, $P_{eq}(z)$, that there is a fraction z of molecules of type 1 in the equilibrium ensemble of states. If $\sigma=0$, the distribution is binomial (determined only by entropy). As σ increases, the peak broadens, and for sufficiently large σ, the distribution is dominated by the two extremes. Although qualitatively similar to Fig. 2, the shapes are not the same and the curve is not well fitted by a beta distribution. One way of achieving this equilibrium distribution in a dynamical simulation would be to consider a ring of fixed size and use moves that exchange one molecule in the ring for a random molecule from the exterior. This equilibrium system would not be a good model of the system of growth and division that we are interested in here. Assemblies in our system do not have a fixed size. There is either steady growth or steady shrinkage, depending on the external concentration of molecules, C. It would be possible to find a specific value of C where the average growth rate is zero, but there would still be large size fluctuations among assemblies. Our system is maintained out of equilibrium by the continual process of addition of new molecules and division of the assemblies. Growth and division are essential parts of this model, and it is not equivalent to the equilibrium model.

Results – Catalytic Model

We will now compare the self assembly model with the catalytic model in which there are also just two types of molecule. We will define a single catalytic rate parameter β and set the catalysis factors $\beta_{11}=\beta_{22}=\beta$ for molecules of the same type, but $\beta_{12}=\beta_{21}=0$ for molecules of different types. In the self-assembly model, it is obvious that molecules will

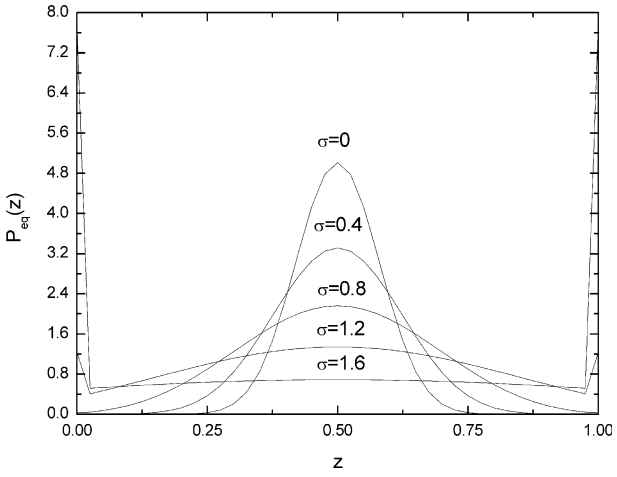

Fig. 6 Equilibrium composition distribution for a ring of size 40

separate into domains of the same type. In the catalytic model, this is less obvious, because although molecules of each type catalyze the addition of further molecules of their own type, they also catalyze their removal. Nevertheless, compositional inheritance does occur in this model and the results are qualitatively similar to the self-assembly model.

We chose parameters N=100, M=40, C=5.0 and D=1.0. Figs. 7 and 8 show $P(z)$ and $\Phi(x)$ in the catalytic model, defined in the same way as for Figs. 2 and 3. Larger values of β correspond to more U-shaped probability distributions and stronger heredity. $P(z)$ becomes U-shaped around β=8, but very high values of β around 300 are required before $\Phi(x)$ becomes U-shaped. High values of β are also required before high heredity values (I close to 1) are obtained. Figure 9 shows I, and the two values of θ as a function of β. The plots of $p(z_{\text{offspring}} \mid z_{\text{parent}})$ for this model are qualitatively similar to those for the self-assembly model and are shown in Fig. 10. The $\Phi(x)$ measurements are well fitted by the beta distribution, as is shown on Fig. 8. The shapes of the $P(z)$ distributions are not well described by the beta distributions; therefore we have not added these fitting curves to Fig. 7.

For both self-assembly and catalytic models, the analogy with the Moran model at the population level is quite strong, because these models are also birth and death processes and we are considering two types of assemblies that are analogous to the two alleles in the Moran model. The analogy is nevertheless not exact, because by counting assemblies as one of just two types, we lose information on the internal composition. The fact that the beta distribution (which is known to be the solution of the Moran model) should fit the data for the $\Phi(x)$ curves in both models shows that the analogy works fairly well. However, the internal composition distribution $P(z)$ depends on the details of the addition and removal steps for the different types of molecule. There is no equivalent of this in the Moran model. Therefore there is no reason why a beta distribution should apply. The beta distribution gives a convenient family of curves symmetric about 1/2, and it happens that these curves are a better description of the $P(z)$ data for the self-assembly case than the catalytic case.

It should be clear that the state of steady growth and division that arises in the catalytic model is very different from thermal equilibrium. Addition and removal rates for each configuration are equal in this model, which is equivalent to setting interaction energies equal for each possible pair of neighbours. Thus, if the system were allowed to go to

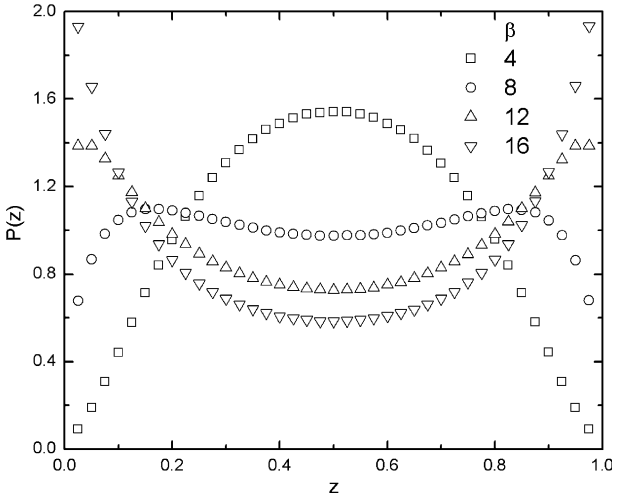

Fig. 7 Catalytic model. Distribution $P(z)$ of the fraction of type 1 molecules in an individual at the point where it reaches maximum size. *Symbols* show simulation results with various values of the catalysis parameter β

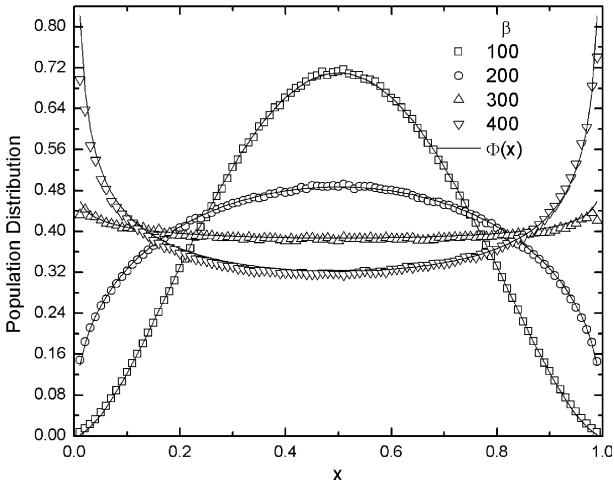

Fig. 8 Catalytic model. Distribution $\Phi(x)$ of the fraction of type 1 individuals in the population. *Symbols* show simulation results with various values of the catlysis parameter β. *Curves* are fits of the beta distribution to the data

equilibrium at a fixed size, then all configurations would have equal probability. Therefore the distribution of compositions would look like the curve for $\sigma=0$ in Fig. 6.

Self-assembly Model with Many Types of Molecule

The results presented above have dealt with the simplest case of only two types of molecule because this is sufficient to demonstrate the principle and because the genetic analogy is easiest to see. However, the stable compositions are uniform in this case. It is already known that the lipid world models of Segré et al. (2000, 2001a, b) stable compositions arise that contain mixtures of molecules. Therefore, in this section, we briefly investigate the self-assembly model in the case where the number of types of molecule, K, is greater than two, and show that stable compositions with more than one type of molecule can arise. There are $K(K+1)/2$ possible nearest neighbour interactions. We assume that the energy ε_{ij}

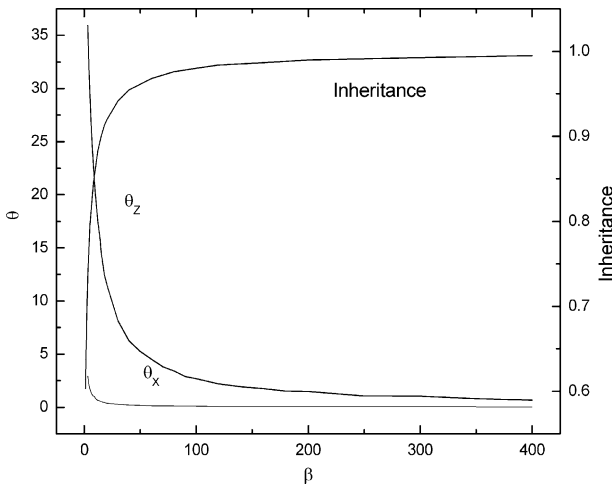

Fig. 9 Catalytic model. *Left hand scale* shows the fitted values of θ for $P(z)$ and $\Phi(x)$ (denoted θ_z and θ_x) as a function of β. *Right hand scale* shows the inheritance function I

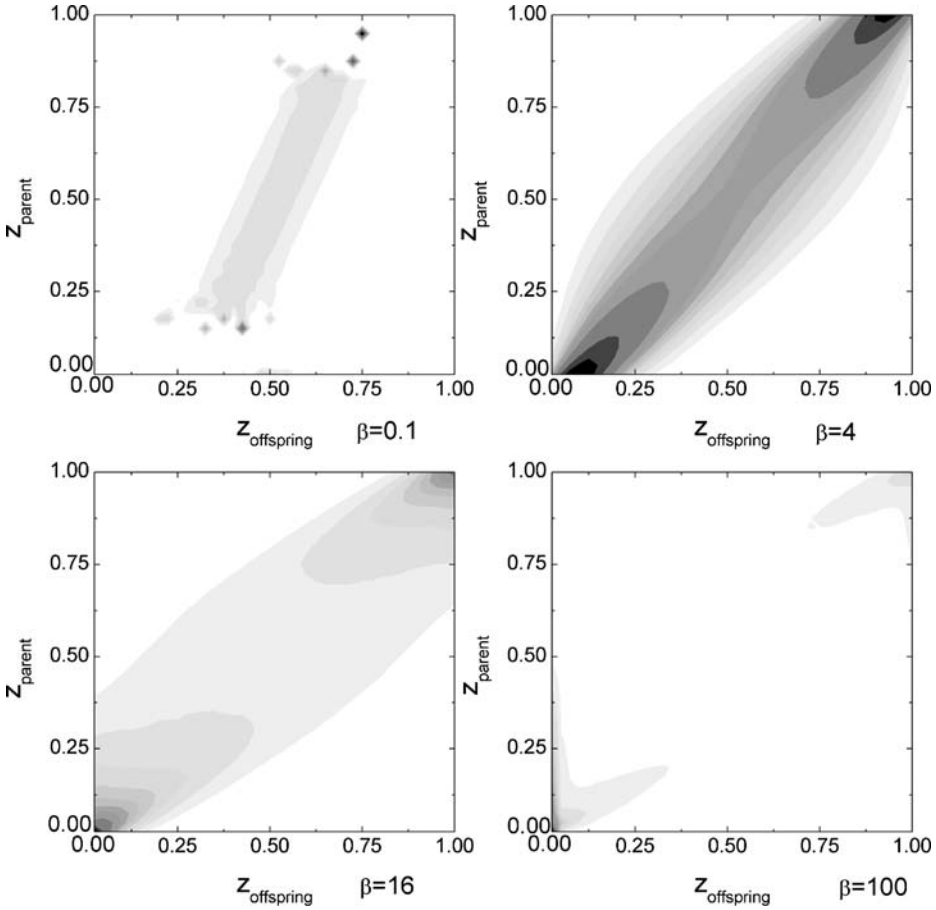

Fig. 10 Catalytic model. Probability distributions of the offspring composition $z_{\text{offspring}}$ conditional on the composition of their parent z_{parent}

of each interaction is chosen randomly from a Gaussian distribution with mean value of 0 and a standard deviation such that $\left\langle \varepsilon_{ij}^2 \right\rangle^{1/2} \Big/ k_{\text{B}}T = \sigma$.

Table 1 shows the values of $\varepsilon_{ij}/k_{\text{B}}T$ in an example of one such random matrix for 10 types of molecules with σ=8. The most negative element on the diagonal is ε_{22} and the most negative off-diagonal element is ε_{18}. The two most stable molecular configurations are the uniform configuration 222222 and the alternating configuration 181818. Figure 11 shows the number of molecules of each type in a simulation that begins with all individuals composed only of twos. Since the population size is N=100, and the maximum assembly size is M=40, the maximum number of molecules is 4,000. The actual number is less than this because there is a distribution of sizes of assemblies that are smaller than M. The number of twos remains high in the population for a long time because most individuals consist of strings of twos. In the early part of the simulation, ones and eights show up as sharp peaks. This is due to the presence of a few individuals with mixed domains such as 18181822222. Later in the simulation, individuals exclusively of 18181818 arise and outcompete the 222222 individuals. The number of twos then falls to a very low level that is

not visible on this scale. The numbers of molecules of types other and 1, 2, and 8 are also too low to be visible on this scale. In both Figs. 11 and 12, the numbers of types 1 and 8 are too close to be distinguishable from one another, which indicates that these molecules almost always occur in blocks of 181818.

Figure 12 shows a simulation with σ=4, such that the values of ε_{ij}/k_BT are half those shown in Table 1. In this case neither of the stable configurations dominates the whole population. Instead, there is a mixture of individuals of each of the two stable configurations and individuals with mixed domains. The frequencies of the other types of molecule are lower than 1, 2, and 8, but are high enough to be visible at this value of σ. Another interesting thing in Fig. 12 is that the 222222 configuration is more frequent than 181818, even though the latter has the lower energy. This is because the 181818 configuration grows more slowly at this value of σ. The alternating configuration has the disadvantage that both ε_{11} and ε_{88} are positive in this random matrix, so inserting either a 1 or an 8 into the alternating configuration is slow. We have observed that growth of the alternating configuration can be aided by defects such as 1818418181, because both ε_{41} and ε_{48} are negative and ones and eights can be easily added next to the defect. Another disadvantage of the alternating configuration is that it can be disrupted by diffusion, i.e. 181818 becomes 188118, and removal of the two molecules that were exchanged then becomes favourable. In the case of 222222, diffusion makes no difference. The relative advantages and disadvantages of different configurations depend on σ. We already saw that 181818 out-competes 222222 at σ=8. The above results confirm that compositional inheritance occurs in the self-assembly model in cases where there are many types of molecule available and that it is possible to have stably inherited compositions that contain more than one type of molecule.

Discussion and Conclusions

Inheritance means the passing on of information from an individual to its descendants. It is well known that this can be done by copying gene sequences, but the lipid world models demonstrate that compositional information can also be passed on without genes. It is pleasing that there should be such a good analogy between the models studied here and the Moran model, which is a classic model in population genetics. This underlines the fact that compositional inheritance really is a form of inheritance.

Table 1 Random matrix of interaction energies with standard deviation σ=8

	1	2	3	4	5	6	7	8	9	10
1	11.34	3.39	−8.39	−6.64	14.70	−5.11	12.76	−25.65	1.26	−2.28
2	3.39	−12.37	−5.34	12.13	23.06	0.40	2.15	−0.09	8.22	−1.83
3	−8.39	−5.34	19.76	−7.24	6.46	−2.02	10.56	1.91	−11.42	−5.95
4	−6.64	12.13	−7.24	3.63	9.42	−8.79	7.37	−8.76	−2.73	−11.19
5	14.70	23.06	6.46	9.42	−1.81	10.62	5.05	−3.34	−5.16	−5.47
6	−5.11	0.40	−2.02	−8.79	10.62	−2.70	−8.49	10.43	8.04	4.80
7	12.76	2.15	10.56	7.37	5.05	−8.49	4.92	−6.57	2.90	−3.74
8	−25.65	−0.09	1.91	−8.76	−3.34	10.43	−6.57	10.95	17.17	1.78
9	1.26	8.22	−11.42	−2.73	−5.16	8.04	2.90	17.17	−6.14	5.96
10	−2.28	−1.83	−5.95	−11.19	−5.47	4.80	−3.74	1.78	5.96	6.20

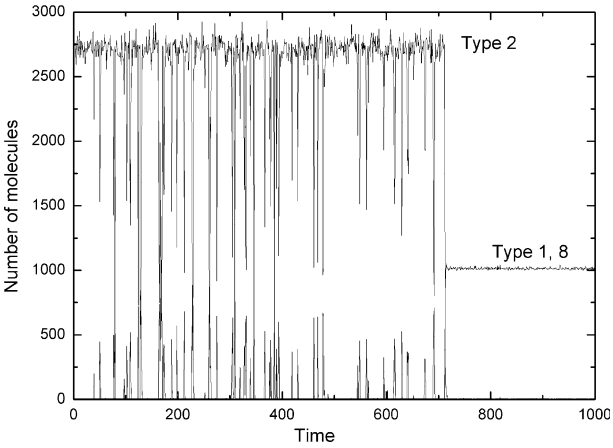

Fig. 11 Self-assembly model with 10 types of molecules and interaction energies as in Table 1. Standard deviation of energies is such that σ=8. The number of molecules of types 1, 2 and 8 are shown as a function of time. Other parameter values are N=100, M=40, C=1.0 and D=1.0. At the beginning of the simulation, each assembly has 20 type 2 molecules

Compositional inheritance has previously been discussed mainly in the context of one particular version of the lipid world model (Segré et al. 2000, 2001a, b). Our object here has been to investigate to what extent the same phenomenon occurs in models that are defined in significantly different ways. We argued that the self-assembly of lipid micelles and vesicles is driven by the thermodynamics of the interactions between them, rather than by catalysis. Therefore we studied a model for self assembly, in which the molecules differ in the energy of the interactions between them, and compared this to a model for catalysis, in which the molecules differ in their catalytic properties. The result is that both models show compositional inheritance, but for different reasons. In the self-assembly model, non-random compositions of molecules arise because they have mutually attractive interactions, whereas in the catalytic model they arise because of mutually favourable catalysis.

In the preceding section on multiple types of molecules, we tested only the self-assembly model and not the catalytic model. We believe that the latter would also show compositional inheritance, as has already been demonstrated in the catalytic models of Segré et al. However, we did not pursue the catalytic model with multiple types because, if the model is supposed to represent very simple molecules like lipids, the kinetic rules of the self-assembly model make much more sense that those of the catalytic model. It is clear that

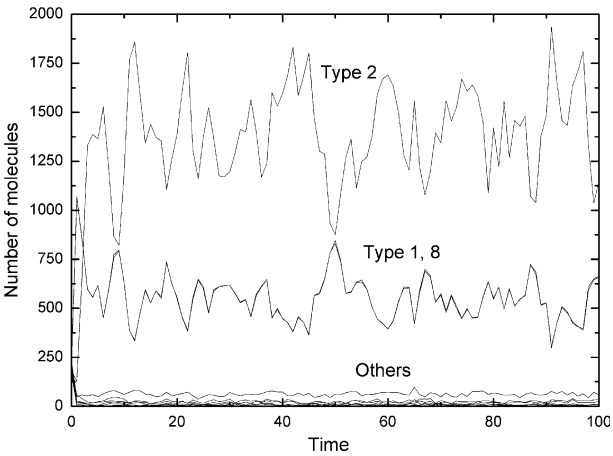

Fig. 12 As Fig. 11 except that σ=4. Low numbers of molecules of types other than 1, 2 and 8 are also visible

interaction energies between neighbouring molecules can differ according to what types of molecules are involved. On the other hand it is not clear that simple lipids have the ability to be very specific catalysts in the way that is required for this model. Values of β in the hundreds are required for the catalytic model to work well. One can imagine highly adapted catalysts, such as membrane proteins, greatly speeding up both the addition and removal of molecules from a cell; however, it seems unlikely that a simple lipid could cause one type of neighbouring molecule to slide into an assembly a hundred times faster than an alternative type and also cause it to slide out a hundred times faster, and yet have identical interaction energy with the two types. Thus, although we have shown that compositional inheritance works for both self-assembly and catalytic models, we feel the self-assembly model is more relevant for assemblies of simple molecules.

Previous lipid world models included no spatial structure. In the self-assembly model, we argued that differences in interaction energies between different types of molecules are sufficient to affect the composition of assemblies. If so, then these differences in interaction energies will be sufficient to cause spatial segregation of molecules. Therefore it seems important to us to account for spatial structure in these models in some way. We used a one-dimensional ring structure because this is very simple to simulate. We presume that compositional inheritance would also occur in models with a two-dimension membrane structure or a three-dimensional globule structure. One dimension is the least favourable case for this kind of model because there is no long-range order. For example, there is no finite temperature phase transition in the one dimensional Ising model, whereas a phase transition does occur in two and three dimensions (Yeomans 1992). Thus, the fact that compositional inheritance does occur in both of our models argues for the generality of the phenomenon, and suggests that compositional inheritance is a concept that has broad relevance beyond the specific example of the model in which it was introduced. Whilst on the subject of spatial arrangement of catalysts, we note that a different point is illustrated by the model of Bradford and Dill (2007), which shows that catalysts that have substrates or products in common tend to associate spatially to form complexes carrying out more than one reaction.

Even though the models we studied demonstrate physical interactions between molecules are sufficient to cause non-random assemblies and to cause inheritance of compositional information, we are still somewhat dissatisfied with the lipid world scenario. It is difficult to see how such assemblies can evolve towards more complex systems unless we incorporate chemical reactions into the model. Dimerization reactions were in fact included in the original GARD model (Segré et al. 1998), but were later removed. Dimers and trimers have been re-added in the latest version (Shenhav et al. 2007). Chemical reactions create a much higher degree of diversity than would be present in the environment, and they lead to the possibility of forming an autocatalytic set of reactions that remains in a non-equilibrium state, like the metabolism of a cell. In the various versions of the GARD model, there are already a large number of monomers and diversity further increases by forming dimers and trimers. In the models for autocatalytic sets of copolymers (Kauffman 1986, 1993; Farmer et al. 1986; Bagley and Farmer 1991), diversity arises because of the exponential number of polymer sequences that can be made from a small number of monomers. This latter case seems more reasonable as a model for either nucleic acids or proteins. Models of copolymer autocatalytic sets have dealt with large deterministic systems and have not so far considered compartments, but it seems reasonable to consider that lipid vesicles may have housed such reaction sets. In this context, we note that Goldstein (2006) has studied another model of populations of assemblies containing complex reaction networks, and finds that compositional inheritance occurs. However, the

rates of increase and decrease in molecular concentrations were taken to be linear functions of the concentrations of the other molecules, and it is not clear that these equations are a good description of the rates of reactions in autocatalytic sets. In general, compositional inheritance would be relevant for vesicles containing autocatalytic sets because the correct mix of chemical reactants would have to be passed on when the vesicle divided. However, it would not be reasonable to call this a lipid world. The molecules involved in the reactions would probably not be lipids, and the role of the lipid might simply be to form the membrane that houses the reactions. One single type of lipid could do this job, and there need be no information or heredity within the membrane itself.

Luisi et al. (1999) and Walde (2006) have emphasized the many roles that surfactant assemblies may have in the origin of life. The conditions under which shape changes leading to budding and division of vesicles can occur have been closely investigated theoretically (Seifert et al. 1991; Bozic and Svetina 2004; Macia and Solé 2007). There has also been a lot of experimental progress towards the construction of systems in which some aspects of living cells are contained within self-assembled lipid vesicles (Pohorille and Deamer 2002; Hancyk and Szostak 2004; Luisi et al. 2006; Solé et al. 2007; Monnard et al. 2007). These experiments demonstrate that lipid vesicles have the right chemical and physical properties to act as environments for early molecular evolution. However, they stop short of demonstrating inheritance of compositional information in the lipids. In fact, we know that modern cells are crucially dependent on information-carrying polymers. It is clear that polymers such as RNA evolved very early in the history of life. Once present, they could be used as catalysts and as genetic material and the possibility of evolution via RNA replication and mutation was then opened up. It is difficult to conceive of evolution beyond the most rudimentary level without the presence of long polymers. Therefore it still seems to us that the crux of the problem of the origin of life is to show how autocatalytic sets involving specific sequences of information-carrying polymers arose. Self-assembled lipid vesicles appear to us as the most likely place in which the chemistry of biopolymer synthesis got going. The possibility that the lipid components of the vesicles showed compositional inheritance in their own right is an interesting one, but the creation of the information-carrying polymers within a vesicle environment would not necessarily require this.

Acknowledgements This work was supported by the Canada Research Chairs organization, NSERC (the Natural Sciences and Engineering Research Council of Canada), and SHARCNET (Shared Hierarchical Academic Research Computing Network).

References

Bachmann PA, Luisi PL, Lang J (1992) Autocatalytic self-replicating micelles as models for prebiotic structures. Nature 357:57–59

Bagley RJ, Farmer JD (1991) Spontaneous emergence of a metabolism. In: Langton CG, Taylor C, Farmer JD, Rasmussen S (eds) Artificial life II, Santa Fe Institute Studies in the Science of Complexity. Addison-Wesley, Reading, MA, pp 93–140

Bartel DP, Unrau PJ (1999) Constructing an RNA world. Trends Biochem Sci 24:M9–M13

Bozic B, Svetina S (2004) A relationship between membrane properties forms the basis of a selectivity mechanism for vesicle reproduction. Eur Biophys J 33:565–571

Bradford JA, Dill KA (2007) Stochastic innovation as a mechanism by which catalysts might self-assemble into chemical reaction networks. Proc Nat Acad Sci USA 104:10098–10103

Dyson FJ (1999) Origins of life, 2nd edn. Cambridge University Press, Cambridge

Ewens WJ (2004) Mathematical population genetics: theoretical introduction. Springer, New York

Farmer JD, Kauffman SA, Packard NH (1986) Autocatalytic replication of polymers. Physica 22D:50–67
Fellerman H, Solé RV (2007) Minimal model of self-replicating nanocells: a physically-embodied, information-free scenario. Phil Trans R Soc B 362:1803–1811
Ganti T (1979) A theory of biochemical supersystems and its application to problems of natural and artificial biogenesis. University Park Press, Baltimore, MD
Gilbert W (1986) Origins of life – the RNA world. Nature 319:618–618
Goldstein RA (2006) Emergent robustness in competition between autocatalytic chemical networks. Orig Life Evol Biosph 36:381–389
Hancyk MM, Szostak JW (2004) Replicating vesicles as models of primitive cell growth and division. Curr Opin Chem Biol 8:660–664
Higgs PG, Pudritz RE (2007) From protoplanetary disks to prebiotic amino acids and the origin of the genetic code. In: Pudritz RE, Higgs PG, Stone J (eds) Planetary systems and the origins of life. Cambridge Series in Astrobiology. vol. 3. Cambridge University Press, Cambridge, pp 62–88
Jeffares DC, Poole AM, Penny D (1998) Relics from the RNA world. J Mol Evol 46:18–36
Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science 292:1319–1325
Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–221
Karlin S, McGregor J (1962) On a genetics model of Moran. Proc Camb Philos Soc 58:299–311
Kauffman SA (1986) Autocatalytic sets of proteins. J Theor Biol 119:1–24
Kauffman SA (1993) Origins of order. Oxford University Press, Oxford
Luisi PL, Walde P, Oberholzer T (1999) Lipid vesicles as possible intermediates in the origin of life. Curr Opin Colloid Interface Sci 4:33–39
Luisi PL, Ferri F, Stano P (2006) Approaches to semi-synthetic minimal cells: a review. Naturwissenschaften 93:1–13
Macia J, Solé RV (2007) Protocell reproduction in a spatially extended metabolism–vesicle system. J Theor Biol 245:400–410
Monnard PA, Luptak A, Deamer DW (2007) Models of primitive cellular life: polymerases and templates in liposomes. Phil Trans R Soc B 362:1741–1750
Moran PAP (1958) Random processes in genetics. Proc Camb Philos Soc 54:60–71
Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Biol 39:99–123
Pohorille A, Deamer D (2002) Artificial cells: prospects for biotechnology. Trends Biotech 20:123–128
Segré D, Lancet D, Kedem O, Pilpel Y (1998) Graded auutocatalysis replication domain (GARD): kinetic analysis of self-replication in autocatalytic sets. Orig Life Evol Biosph 28:510–514
Segré D, Ben-Eli D, Lancet D (2000) Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies. Proc Nat Acad Sci USA 97:4112–4117
Segré D, Shenhav B, Kafri R, Lancet D (2001a) The molecular roots of compositional inheritance. J Theor Biol 213:481–491
Segré D, Ben-Eli D, Deamer DW, Lancet D (2001b) The lipid world. Orig Life Evol Biosph 31:119–145
Seifert U, Berndl K, Lipowski R (1991) Shape transformations of vesicles: phase diagram for spontaneous curvature and bilayer-coupling models. Phys Rev A 44:1182–1202
Shapiro R (1988) Prebiotic ribose synthesis – a critical review. Orig Life Evol Biosph 18:71–85
Shapiro R (2006) Small molecule interactions were central to the origin of life. Quart Rev Biol 81:105–125
Shenhav B, Oz A, Lancet D (2007) Coevolution of compositional protocells and their environment. Phil Trans R Soc B 362:1813–1819
Solé RV, Munteanu A, Rodriguez-Caso C, Macia J (2007) Synthetic protocell biology: from reproduction to computation. Phil Trans R Soc B 362:1727–1739
Szabo P, Scheuring I, Czaran T, Szathmary E (2002) In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. Nature 420:340–343
Szathmary E, Santos M, Fernando C (2005) Evolutionary potential and requirements for minimal protocells. Top Curr Chem 259:167–211
Walde P (2006) Surfactant assemblies and their various possible roles for the origin of life. Orig Life Evol Biosph 36:109–150
Yeomans JM (1992) Statistical mechanics of phase transitions. Oxford University Press, Oxford
Zaher HS, Unrau PJ (2007) Selection of an improved RNA polymerase with superior extension and fidelity. RNA 13:1017–1026