Coexistence and error propagation in pre-biotic vesicle models: A group selection approach

José F. Fontanari\textsuperscript{a,*}, Mauro Santos\textsuperscript{b,c}, Eörs Szathmáry\textsuperscript{c,d}

\textsuperscript{a}Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos SP, Brazil
\textsuperscript{b}Department of Genética i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
\textsuperscript{c}Collegium Budapest, Institute for Advanced Study, Szenthéromság u. 2, H-1014 Budapest, Hungary
\textsuperscript{d}Department of Plant Taxonomy and Ecology, Eötvös University, 2 Ludovika tér, H-1083 Budapest, Hungary

Received 26 April 2005; received in revised form 29 June 2005; accepted 23 July 2005
Available online 21 October 2005

Abstract

Compartmentalization of unlinked, competing templates is widely accepted as a necessary step towards the evolution of complex organisms. However, preservation of information by templates confined to isolated vesicles of finite size faces much harder obstacles than by free templates: random drift allied to mutation pressure wipe out any template that does not replicate perfectly, no matter how small the error probability might be. In addition, drift alone hinders the coexistence of distinct templates in a same compartment. Here, we investigate the conditions for group selection to prevail over drift and mutation and hence to guarantee the maintenance and coexistence of distinct templates in a vesicle. Group selection is implemented through a vesicle survival probability that depends on the template composition. By considering the limit case of an infinite number of vesicles, each one carrying a finite number of templates, we were able to derive a set of recursion equations for the frequencies of vesicles with different template compositions. Numerical iteration of these recursions allows the exact characterization of the steady state of the vesicle population—a quasispecies of vesicles—thus revealing the values of the mutation and group selection intensities for which template coexistence is possible. Within the main assumption of the model—a fixed, finite or infinite, number of vesicles—we find no fundamental impediment to the coexistence of an arbitrary number of template types with the same replication rate inside a vesicle, except of course for the vesicle capacity. Group selection in the form of vesicle selection is a must for compartmentalized primordial genetic systems even in the absence of intra-genomic competition of different templates.

Keywords: Pre-biotic evolution; Group selection; Error threshold; Package models; Compartmentalization

1. Introduction

The apparently harmless observation that the length of a replicating polymer (i.e. RNA-like template) is limited by the replication accuracy per nucleotide has been a challenge to the development of the field of pre-biotic evolution for more than three decades (Eigen, 1971). Within the quasispecies framework, which may be used as a paradigm in this discussion, templates have to replicate with high accuracy in order to reach a certain length, a requirement that seems very unlikely to be met without the aid of specialized catalysts (peptide enzymes). To build those catalysts, however, a blueprint is needed that amounts to a large nucleotide sequence, which itself cannot be maintained without the catalysts. This is the so-called Catch-22 of the origin of life: no large genome without enzymes, and no enzymes without a large genome (Maynard Smith, 1983). The information crisis in pre-biotic evolution originates from this observation together with the finding that, except in a trivially degenerate case, two or more templates which differ significantly from each other cannot coexist (Swetina and Schuster, 1982). These two simple results have struck a fatal blow on the simplistic view of the emergence of a complex genome from a collection of
competing templates: additional assumptions regarding cooperation between templates are mandatory to overcome this impasse.

Most proposals to circumvent the information crisis have conformed to the limitation of information coded by a single template, focusing instead on the conditions that guarantee the stable coexistence of different templates (for rare exceptions, see Scheuring, 2000; Szabó et al., 2002). Since the total information content is the product of the number of different templates and the maximum information coded per template, provided the template types have the same concentration, ensuring the coexistence of different templates resolves the information crisis. In this vein, Eigen (1971) and Eigen and Schuster (1979) proposed the hypercycle, i.e. a catalytic feedback network in which each template helps in the replication of the next one, in a secondary cycle closing on itself (second-order autocatalysis). Originally, the enzymatic function (i.e. replicate activity) was thought to be carried out by the encoded proteins (the “realistic” hypercycle: Eigen and Schuster, 1979), but recently also the hypercycle is projected into the RNA world (Gilbert, 1986), with RNA molecules acting as templates as well as enzymes (cf. Zintzaras et al., 2002). This view gained plausibility when the ability of polynucleotides to exert various catalytic (enzymatic) functions was discovered (see Doudna and Szostak, 1989; Doudna and Lorsch, 2005 for review). However, if the number of hypercycle members is greater than four, the template concentrations will vary with time (Eigen and Schuster, 1979), periodically decreasing to very small values, thus making large hypercycles vulnerable to extinction via fluctuations (Nuño and Tarazona, 1994). The information gain due to the coexistence of different templates in the hypercycle is then not significant. In addition, the hypercycles have attracted criticisms for other reason: as first pointed out by Maynard Smith (1979), giving catalytic support in such molecular networks is in fact an altruistic behaviour and so hypercycles are easy targets to parasites, i.e. molecules that do not reciprocate the catalytic support they receive. To solve the conundrum of dynamic coexistence of different templates Szathmáry and Demeter (1987), elaborating on the package model proposed by Niesert et al. (1981), described the “stochastic corrector model” as an effective treatment against parasites and of the conflict among genes due to internal competition; namely, the compartmentalization or packaging of unlinked templates that are replicated by genes with a non-specific replicate functionality in isolated vesicles that reproduce by clonal selection (see also Grey et al., 1995), so that infected vesicles simply die out preventing the spreading of the mutant parasites to the healthy ones (but see Santos et al., 2003).

Since the interaction of templates within vesicles is essentially competitive and the vesicle dynamics, usually fission, would eventually lead to the loss of essential genes for survivorship (i.e. the assortment load), some sort of selection at the vesicle or compartment level is obviously required for these models as an alternative to hypercycles (see, e.g. Niesert et al., 1981; Szathmáry and Demeter, 1987; Alves et al., 2001; Zintzaras et al., 2002; Hogeweg and Takeuchi, 2003). However, although the coexistence of a few distinct templates has been demonstrated for sure, a common criticism is that the package model is a very sloppy system of information integration because the major difficulty for conserving a complete set of genes is the growth difference between replicators (e.g. Suzuki and Ono, 2003). Thus, in the model of Niesert et al. (1981), the number of different templates per compartment cannot exceed three given a realistic parameter setting. Therefore, although conceptually simpler than the hypercycles it is not clear whether the information content of the package models could be substantially larger than the information content in the free-template case. Similar to the package model of Niesert et al. (1981) and the stochastic corrector model (Szathmáry and Demeter, 1987) we assume here that the templates are replicated by a non-specific replicate. How this replicate evolved is a serious problem we leave open as there is no currently known metabolic path to RNA in pre-biotic settings, and the attainment of a fully functional RNA replicate is a very difficult enterprise (Johnston et al., 2001). Also, we do not know how many different templates (genes) can coexist in the packages.

In comparison with the hypercycle, very little is known about the dynamics and steady states of package models, since their greater complexity frustrates any attempt to carry out a thorough analysis of the space of parameters that determine the evolution of templates and vesicles. To bypass this difficulty, here we study a package/compartment model using the discrete-time mathematical formalism of group selection (Eshel, 1972; Levin and Kelmer, 1974; Aoki, 1982; Silva and Fontanari, 1999) in which a countable infinity of groups (vesicles, in the case) are isolated from each other (see, e.g. Wade, 1978; Boorman and Levitt, 1980 for reviews). The total number of templates confined in a vesicle is finite and is the same for all vesicles. Because of random genetic drift, this feature implies that in the absence of group selection (differential extinction depending on the internal template composition of the group) only one type of template will be present in a given vesicle in the equilibrium regime. Of course, the surviving template type may not be the same in different vesicles. Hence, by barring the coexistence of distinct templates in the same vesicle, random drift plays a similar role as the template assortment procedure of the stochastic corrector model. Here, a group is viewed as an isolated set of templates that maintains its integrity during its entire existence. This view contrasts with the ephemeral groups of Wilson’s (1980) formulation widely used in pre-biotic evolution (Michod, 1983; Donato et al., 1997; Alves et al., 2001) and viral dynamics (Szathmáry, 1992) in which the group structure is dissolved each generation to form a global mating pool; but comes close to the stochastic corrector model. In particular, in the transient “group” formulation random drift plays no role at all,
since different templates in distinct vesicles are likely to be assigned to the same vesicle during the group re-assembling procedure after the mating stage. Our approach allows us to derive exact recursion equations for the frequency of vesicles with a given template composition, which are then easily solved numerically.

In addition to establishing the conditions that must be met to guarantee the coexistence of a given number of templates inside a vesicle, in this contribution we consider explicitly the existence of a special class of templates, so-called error-tail, that appears as a consequence of the replication errors of the original templates and has no effect whatsoever on the survival of the vesicle. Back-mutations and mutations between the original classes of templates are neglected. This simplified formulation facilitated greatly the analysis of the error threshold in Eigen’s model (Maynard Smith, 1983; Scheuring, 2000) as well as in the hypercycle (García-Tejedor et al., 1987; Nuño et al., 1993; Campos et al., 2000). An interesting detailed study of the influence of vesicle dynamics on the information threshold in the case when vesicles are occupied by a single uncorrupted template type and its error tail is given by Hugweg and Takeuchi (2003) in the context of a cellular automaton model, both at the level of templates and vesicles.

### 2. The model

The population is composed of an infinite number of vesicles, each of which encloses exactly $N$ templates. There are $d + 1$ distinct types of templates, labelled by the integers $l = 0, 1, \ldots, d$, where type $l = 0$ denotes templates belonging to the error tail. The replication rates are denoted by $s_0, s_1, \ldots, s_d$ with $s_1 > s_2 > \cdots > s_d > s_0$ without loss of generality. Due to imperfect replication, templates of type $l > 0$ mutate to templates of type $l = 0$ (error tail) with probability $u$. Back-mutations and mutations between uncorrupted templates are neglected. The vesicles are identified according to their template composition by the vector $\hat{k} = (k_0, k_1, \ldots, k_d)$, where the component $k_l$ stands for the number of templates of type $l$ in the vesicle. Since $\sum_{l=0}^{d} k_l = N$, there are $\binom{d + N}{N}$ different types of vesicles. The state of the population at a given generation $t$ is completely specified by the frequencies of vesicles of type $\hat{k}$, which we denote by $Y_t(\hat{k})$ with $\sum_{\hat{k}} Y_t(\hat{k}) = 1$ for all $t$. The life cycle (i.e. one generation) consists of four events—extinction, recolonization, replication and mutation—which take place in this order and are described as follows.

#### 2.1. Extinction

The basic idea of addressing the information crisis problem by ensuring the coexistence of distinct templates is that the information content of a vesicle is split into $d$ parts and all of them must be present at all times to guarantee the viability of the system. In addition, it is desirable that the different templates contribute with approximately the same number to the vesicle composition. Hence, we define the probability that a vesicle of type $\hat{k}$ survives extinction, denoted by $z(\hat{k}) \in [0, 1]$, as a function proportional to the geometric mean of the number of templates confined in the vesicle:

$$z(\hat{k}) = 1 - g + g \frac{d}{N} \left( k_1 k_2 \cdots k_d \right)^{1/d},$$

where $g \in [0, 1]$ is a parameter measuring the intensity of the group selection pressure or the benefit accrued to the vesicle from the presence of the templates within it. Survival is guaranteed (i.e. $z = 1$) only for vesicles which have the even template composition $k_l = N/d$ for $l = 1, \ldots, d$ regardless of the value of $g$, provided this value is not zero. Vesicles that lack one or more template types are assigned the baseline survival probability $1 - g$, so that the selective pressure against these vesicles increases with increasing the value of $g$. In this sense, this formulation is akin to that of the single sharp-peak replication landscape of the quasispecies model (Eigen, 1971). The geometric mean that appears in Eq. (1) has been repeatedly used to model the metabolic function associated in recent models of replicator networks (e.g. Czarán and Szathmáry, 2000). The baseline survival probability $1 - g$, however, is a new ingredient of our model that takes into account the experimental finding that vesicles actually grow and divide even in the absence of templates (Hanczyc et al., 2003). We note that whenever a mutation occurs a useful template, say, template 1 is replaced by an inert template of the error-tail class, $k_1$ decreases and so does the survival probability of the vesicle, since the numbers of the other uncorrupted template $k_2, \ldots, k_d$ are unaltered. Hence, mutations have a direct deleterious effect on the protocell replication rate.

#### 2.2. Recolonization

As the result of extinction, a fraction $1 - \sum_{\hat{k}} Y_t(\hat{k})$ of vesicles disappears and is then replaced or recolonized by the surviving vesicles. This is done by reproducing these vesicles in proportion to their frequencies in the population, resulting in the following new vesicle frequencies:

$$\frac{z(\hat{k}) Y_t(\hat{k})}{\sum_{\hat{j}} z(\hat{j}) Y_t(\hat{j})}$$

from which it becomes clear that the recolonization procedure introduces an effective competition between the vesicles. In fact, Eq. (2) prompts the interpretation of the parameter $z(\hat{k})$ as the replication rate of a vesicle of composition $\hat{k} = k_0, \ldots, k_d$. We note, however, that the procedure used here for recolonization (Eshel, 1972; Levin and Kilmer, 1974; Aoki, 1982; Silva and Fontanari, 1999) implies the instantaneous replacement of the extinct vesicles by the surviving ones with the probability given
in Eq. (2). For example, in an extreme situation, in which only one vesicle passes the extinction stage, this sole vesicle will replenish the entire population (infinite or finite) in a single time step. In any event, provided that \( g \) does not equal 1, we can safely ignore this difficulty.

2.3. Replication

This process describes the replication of the templates inside the vesicles. As usual, we assume that the number of offspring that a template contributes to the new generation is proportional to its relative replication rate. Since only \( N \) offspring are chosen to replace the \( N \) parents we can write the probability that a vesicle of type \( k \) changes to a vesicle of type \( j \) as

\[
R(j|k) = \frac{N!}{j_0!j_1! \ldots j_d!} w_0^{i_0} w_1^{i_1} \ldots w_d^{i_d},
\]

where \( w_l = k_l s_l / \bar{w} \) for \( l = 0, 1, \ldots, d \) are the relative replication rates of the template types. Here, \( \bar{w} = k_0 s_0 + k_1 s_1 + \cdots + k_d s_d \) is the average replication rate of the entire population.

2.3.1. Mutation

The effect of the mutation process is solely to increase the number of templates in the error tail. Hence, the probability that a vesicle of type \( j \) changes to one of type \( i \) is simply

\[
M(i|j) = \begin{pmatrix} i \end{pmatrix} d^{i-i} (1-u)^i \prod_{l=1}^d d^{j_l-i_l} (1-u)^j_l
\]

with \( j_l > i_l \) for \( l = 1, \ldots, d \) and \( j_0 \leq i_0 \).

The processes of template replication and mutation can be described by a single transition matrix \( T(|k) = \sum_j M(|j)R(|j) \), that gives the probability that a vesicle of type \( k \) changes to a vesicle of type \( i \), due to replication immediately followed by mutation of the offspring. In fact, the present approach is feasible only because we can carry out the sum over the intermediate states \( j \) and derive a closed expression for this combined transition matrix. We find

\[
T(|k) = \frac{N!}{k_0!i_1! \ldots i_d!} \left[ u + w_0 (1-u)^{i_0} \right]
\]

\[
\times \left[ w_1 (1-u)^{i_1} \ldots w_d (1-u)^{i_d} \right],
\]

where \( w_l \) are the relative replication rates given before.

Finally, given the events comprising the life cycle of the templates and vesicles we can immediately write down a set of coupled recursion equations for the frequencies of the different types of vesicles in the population:

\[
Y_{t+1}(|i) = \frac{\sum_j T(|j)z(j)Y_t(|k))}{\sum_k z(k)Y_t(|k))}.
\]

The main difficulty to iterate numerically these equations is that one has to keep track of the order of \( N^d \) frequencies at each generation, thus setting limits to the vesicle capacity \( N \) and the template diversity \( d \) we can investigate. All results presented in this paper are obtained in the stationary regime \( Y_{t+1} = Y_t = Y \). We found these steady states remarkably insensitive to the choice of the initial distribution of templates in the vesicles \( Y_{t=0}(|k) \): provided that there are vesicles that contain all uncorrupted template types, the dynamics will always approach the same steady state. In what follows, we will focus mainly on the fraction \( \Omega_m \) of vesicles that carry copies of \( m \leq d \) uncorrupted template types, regardless of their redundancy.

3. Coexistence of error-free templates

Before considering the effect of imperfect replication and the consequent appearance of the error tail, in this section we focus on the somewhat simpler though fundamental problem of the coexistence of templates with distinct replication rates. By setting \( u = 0 \) and \( Y_{t=0}(|k) = 0 \), if \( k_0 \neq 0 \), our mathematical formalism can be readily used to study this problem as well. As already pointed out, random drift will bar the coexistence of two or more distinct templates in a vesicle of finite capacity. On the other side, competition between templates will lead to the dominance of the template with the highest replication rate in the case of very large vesicles. Here, we address the issue of how strong the group selection pressure must be in order to counterbalance the effects of drift and competition.

Fig. 1, which displays the fraction \( \Omega_2 \) as a function of the group selection intensity \( g \) for \( d = 2 \), \( N = 50 \) and several values of the ratio \( s_2/s_1 \), reveals the existence of a threshold \( s_2/s_1 = 0.25 \) below which random drift and template competition are the dominant forces, resulting then in the complete frustration of template coexistence. To gauge and even disentangle the contributions of drift and competition we can vary the vesicle capacity \( N \). This is illustrated in

![Fig. 1](image-url)
Fig. 2. Critical value of the group selection pressure \( g_c \) as a function of the vesicle capacity \( N \) for \( d = 2, s_2/s_1 = 1.0, 0.75, 0.5 \) and 0.25 in the case of error-free replication \( u = 0 \). Coexistence of the two template types in a same vesicle is guaranteed above \( g_c \). The line for \( s_2/s_1 = 1.0 \) is the fitting \( g_c = 1/N \), the other lines are guides to eye.

Fig. 3. Fraction of vesicles with \( d = 2, 3 \) and 4 different types of templates in their composition \( \Omega_d \) as a function of the group selection pressure \( g \) for \( N = 24 \) in the case of error-free replication \( u = 0 \). The solid lines are the results for degenerate replication rates and the dashed lines are the results for \( s_2/s_1 = 0.8 \) (\( d = 2 \)); \( s_2/s_1 = 0.8, s_3/s_1 = 0.6 \) (\( d = 3 \)); and \( s_2/s_1 = 0.8, s_3/s_1 = 0.6, s_4/s_1 = 0.4 \) (\( d = 4 \)).

Fig. 2 shows which dependence the threshold value \( g_c \), below which coexistence is impossible, on the vesicle capacity \( N \). In the absence of competition \( (s_2/s_1 = 1) \), coexistence is facilitated by increasing the vesicle capacity since this attenuates the effect of drift. The data for this case (lower curve in Fig. 2) are fitted perfectly by the function \( g_c = 1/N \). The situation is more complicated when competition \( (s_2/s_1 < 1) \) is taken into account: after an initial decrease, due to reduction of drift, \( g_c \) starts to increase smoothly towards the singular value \( g_c = 1 \) as the vesicle size grows and consequently template competition is enhanced. As has already been said, our formulation breaks down if the group selection pressure is set to its maximum value, since a singular situation in which all vesicles are extinct at the same time (i.e. \( x_i = 0 \) for all \( i \) such that \( Y_i \neq 0 \)) is likely to arise. In fact, this is exactly what happens in the deterministic regime \( (N \to \infty) \). Since the random fluctuations that promote the diversity of the vesicles are suppressed in this regime, the vesicles soon become all alike and so group selection is turned off. Template competition is then free to act ending in the fixation of the fitter template in all vesicles regardless of the value of \( g < 1 \). Note that the singularity, i.e. the simultaneous vanishing of the numerator and denominator in Eq. (6), occurs at \( g = 1 \) only. Hence, except for the degenerate case \( s_2/s_1 = 1 \), template coexistence is impossible in the deterministic regime.

The rich trade-off between drift and competition is responsible for the appearance of a minimum in the curves \( g_c \) vs. \( N \). This minimum disappears for small values of the ratio \( s_2/s_1 \) indicating that the effect of drift becomes negligible compared to that of competition. However, in the probably more realistic case that the template replication rates are almost degenerate \( (s_2/s_1 \approx 1) \), it is the effect of template competition that is negligible, provided the vesicle capacity is not too large. Fig. 3 illustrates the dependence of \( \Omega_d \) on \( g \) for a vesicle of fixed size \( N = 24 \) and \( d = 2, 3 \) and 4 different template types in the degenerate case (i.e. \( s_1 = s_2 = \cdots = s_d \)) as well as in the more general situation that all replication rates are different. Regardless of whether competition is allowed or not, coexistence of many distinct templates can be achieved by increasing the group selection pressure, though as mentioned before, care must be taken to avoid the singular value \( g = 1 \). In fact, we found no evidence of a threshold on the number of templates beyond which coexistence is precluded. For instance, in the extreme case of \( d = N \) degenerate templates, the recursion equations (6) can be solved analytically, resulting in the simple expression:

\[
\Omega_{d=N} = \frac{1}{g} \left[ g - \left( 1 - \frac{d^1}{d^d} \right) \right],
\]

valid for \( g \gg 1 - d! / d^d \). We note that in this region, \( \Omega_{d=N} \) is a very small quantity, on the order of \( \exp(-d) \) for large \( d \). So there is no fundamental impediment to the coexistence of an arbitrary number \( d \ll N \) of distinct templates in a vesicle of capacity \( N \), provided that the group selection parameter can be set arbitrarily close, but not equal, to 1. Of course, as \( g \) approaches 1 the fraction of vesicles that contain all template types decreases very rapidly and so the vast majority of the vesicles present in the population will lack one or more types. Hence, group selection alone cannot explain the prevalence of that “cooperative” kind of vesicle and so one needs to invoke another mechanism to suppress template competition, which will then permit the rare representatives of the cooperative vesicles (we have proven that group selection can maintain a few vesicles of that kind) to take over the vesicle population. The ingredient missing is probably a selective pressure for
linkage of all templates on a single chromosome (Maynard Smith and Szathmáry, 1993; Santos, 1998). In that sense, the (partial) failure of group selection is welcome, otherwise it would be difficult to explain gene linkage, and hence the origin of chromosomes, if an arbitrary number of unlinked templates could easily be made to coexist within compartments.

To verify how our conclusions are affected by the assumption that there are a countable infinity of vesicles, here we carry out a simple Monte-Carlo simulation to study the situation in which the number of vesicle is fixed to a finite value $M$. Explicitly, we re-interpret Eq. (2) as the relative fitness of vesicles composed of $k_1, k_2, \ldots, k_d$ template types and then use a transition probability analogous to Eq. (3) to generate the new vesicle population from the previous one. Essentially, the same stochastic procedure used to model the template replication inside each vesicle is used to model vesicle replication. The results for $d = 2$ degenerate ($s_2/s_1 = 1$) templates fenced in $M = 10, 20$ and 40 vesicles of capacity $N = 20$ are depicted in Fig. 4 together with the analytical result for $M \to \infty$. The data represent averages over 100 independent runs. The point to be noted here is that the behaviour pattern observed in the case of a finite number of vesicles does not differ qualitatively from the pattern predicted by the deterministic theory: the sole effect of finite $M$ is to increase the value of the threshold $g_c$ so as to make coexistence more difficult. As expected, the agreement between our analytical results and the Monte-Carlo simulations improves as $M$ increases and, in particular, they become practically indistinguishable for $M > 500$. Note, however, that by fixing the number of vesicles we have ruled out the possibility of extinction of the vesicle population, a realistic ingredient that might pose a serious obstacle to template coexistence when $g$ is set too close to 1. Nevertheless, our results apply provided the vesicle population is not extinct, since then it is likely to be very large.

4. The effect of the error tail

The consequences of imperfect replication are catastrophic for templates isolated in compartments of finite capacity: even for a vanishing small mutation probability ($u \approx 0$) and replication rate ($s_0/s_l \approx 0, l = 1, \ldots, d$), the error tail will be fixed in the compartment. In fact, though rare, mutations are recurrent and so they will prevent the fixation of the uncorrupted templates. In addition, the fixation of the error trail is very unlikely to occur because of its small replication rate, but once it does the game is over for the uncorrupted templates. A possible way out of this conundrum is to invoke again group selection—the selection at the vesicle level. The pressing question of how strong group selection must be in order to ensure the survival of templates confined in a vesicle is addressed in this section, where we will concentrate on the interplay between the group selection pressure and the mutation probability in the case of $d = 1$ and 2 templates. Since in the previous section we have shown that introducing competition between templates does not alter qualitatively the behaviour pattern observed in the case of degenerate templates, henceforth we will consider the situation in which all $d$ uncorrupted templates have the same replication rate, which we set to 1 without loss of generality, while templates in the error tail have replication rate $1 - \tau$ with $\tau \in [0, 1]$.

First we consider the case of a single uncorrupted template type ($d = 1$) in a vesicle of capacity $N$. Error-tail templates are not included in the initial assignment of templates to vesicles. In Fig. 5, we show the steady-state frequencies of vesicles with several template compositions. All these frequencies vanish at a critical value of the

![Fig. 4](image1.png)

Fig. 4. Fraction of vesicles with two degenerate types of templates in their composition $\Omega_2$ as a function of the group selection pressure $g$ for $M = 10, 20$ and 40 vesicles of fixed capacity $N = 20$. The solid line is the analytical result for $M \to \infty$.

![Fig. 5](image2.png)

Fig. 5. Steady-state frequencies of vesicles with $i = 0, 4, 8, 12, 16$ and 20 copies of the template as a function of the mutation probability $u$ for $d = 1$, $N = 20$, $\tau = 0.1$ and $g = 0.5$. We note that $Y(20) \to 1$ for $u \to 0$ and $Y(0) \to 1$ for $u \to u_c \approx 0.26$. 
mutation probability \( u_c \approx 0.26 \), except for \( Y(0) \) that tends to 1, indicating that the error tail has become fixed in all vesicles. This figure endorses the portrait of package models in equilibrium as a quasispecies of compartments with internal competition, as demonstrated for the stochastic corrector model (Szathmáry and Demeter, 1987) and, in addition, demonstrates the existence of an error threshold \( u_c \) beyond which the uncorrupted template is lost from the population. A better characterization of this threshold is given in Fig. 6 that shows the fraction of vesicles that contain at least one copy of the uncorrupted template \( \Omega_1 \) as a function of the mutation probability for several values of the group selection pressure. The point here is to emphasize that \( u_c \) goes to zero when the group selection pressure is turned off, in agreement with the qualitative reasoning presented in the beginning of the section. The effect of changing the vesicle capacity \( N \), depicted in Fig. 7, is quite complex and somewhat puzzling. For small \( u \), increasing \( N \) favours the persistence of the uncorrupted template as expected since the effect of drift is attenuated in this case. However, for large \( u \) the situation is reversed, resulting that low capacity vesicles become able to tolerate much higher mutation rates. The reason for that is as follows. According to Eq. (1), for large \( N \) the survival probability is significant only for those vesicles with a large number of templates. Vesicles with few templates have survival probability only slightly larger (order of 1/\( N \)) than the baseline. Thus, by diminishing the number of uncorrupted templates, large mutation rates will weaken the efficacy of group selection to maintain the templates in large vesicles. We note that in the deterministic regime \( \Omega_1 \) becomes the step function \( \Omega_1 = 1 \) for \( u < \tau \) and \( \Omega_1 = 0 \) otherwise. These results are summarized in Fig. 8 where the critical mutation probability is presented against the group selection pressure. For \( g = 0 \) we find \( u_c = 0 \) regardless of the values of \( \tau \) and \( N \); however, this critical value jumps to a finite value for any arbitrarily small value of \( g \) for \( \tau \) not too small. Although our numerical results cannot guarantee the existence of a real discontinuity at \( g = 0 \), at least the extreme sensitivity of the threshold to a non-zero value of \( g \) is clearly demonstrated.

We turn now to the analysis of the case of two uncorrupted template types (\( d = 2 \)) confined in vesicles of finite capacity. As before, error-tail templates are left out of the initial assignment of templates to vesicles. Fig. 9 illustrates the typical equilibrium behaviour pattern of the fraction of vesicles of capacity \( N = 20 \) containing \( m = 0, 1 \) and 2 uncorrupted template types (regardless of their numbers). The group selection pressure, set to \( g = 0.5 \), can maintain the coexistence of the two template types in more than 90% of the vesicles in the absence of replication errors, \( u = 0 \). Initially, increasing the mutation probability (up to \( u = 0.05 \)) results in an increase of vesicles carrying only one type of uncorrupted template (\( m = 1 \)) at the
Vesicles carrying only error-tail templates ($m = 0$) become common for larger values of the mutation probability. The fact that both fractions $\Omega_1$ and $\Omega_2$ vanish simultaneously at a single threshold value is easily understood. According to Eq. (1), vesicles lacking one or more uncorrupted template type are assigned the baseline survival probability, and so group selection is turned off when the vesicles with $m = d$ uncorrupted template types are extinct. Since we have already argued that in the absence of group selection uncorrupted templates cannot be maintained in vesicles of finite capacity, the vesicles with $m < d$ must die out too. In fact, the dependence of the critical mutation probability on the group selection pressure, depicted in Fig. 10, reinforces this point (though for a rather low vesicle capacity $N = 6$), since for $g < 0.18$ group selection is overcome by drift (see Fig. 2), the coexistence of the two uncorrupted template types is impossible and so they are lost in the presence of a vanishing small mutation probability.

5. Conclusion

Consider the standard scenario of pre-biotic evolution: a population of templates (or even a more elaborate system of coupled templates, such as the hypercycle) whose imperfect replication accuracy produces inefficient and useless templates, the error tail, with back-mutations not allowed, as usual. Then confine this population to a vesicle of finite capacity, an often quoted “next” step (cf. Maynard Smith and Szathmáry, 1995) towards the evolution of more complex organisms. The stage for a catastrophe is set, since classical population genetics tells us that sooner or later the useful templates will be lost due to random genetic drift—a fate the templates in the error tail are spared due to the rare but recurrent mutation events. The solution to this conundrum is to admit that the vesicles have a dynamics of their own, which may be affected by their particular template compositions, resulting in a new selection pressure—group selection. This view accords with a setting where metabolic life (i.e. a population of vesicles capable of growth and reproduction) thrived before the appearance of templates, i.e. the metabolism-first hypothesis. As repeatedly pointed out in the literature (see, e.g. Dyson, 1985; Gánti, 1987, 2003), heredity and, consequently, Darwinian evolution are not necessary and must not be invoked at this metabolic stage; although it is not clear at all what perils are in store for such hypothetical populations of metabolic vesicles.

In this paper, we investigated quantitatively the strength of the group selection necessary to counterbalance the intra-group disintegration pressures of mutation, competition and drift, and so guarantee the coexistence of a few different template types in a same vesicle. The formalism employed here builds on the work of Aoki (1982) (see also Boorman and Levitt, 1980; Eshel, 1972; Levin and Kilmer, 1974) on the evolution and maintenance of altruistic traits. In particular, it is assumed that there are an infinite number of vesicles, each of which occupied by a finite number of template types. This assumption allows us to write a set of recursion equations for the frequencies of the vesicle types in the population. Due to the large combinatorial number of all possible vesicle compositions, this approach is feasible for at most $d = 4$ template types and vesicle capacity $N < 100$. We note, however, that most quantitative studies of package models have concentrated on the coexistence of only two template types. The advantage of this approach is that we can fully explore the space of the control parameters (except for $d$ and $N$) and so determine with certainty the critical values of the parameters that delimitate the regions in that space where template coexistence is possible or not. Rather surprisingly, we find no fundamental impediment to the coexistence of an arbitrary number of template types $d \leq N$, provided the group selection pressure is sufficiently strong. This is in stark contrast with the findings for the hypercycle, for which no matter the strength of the kinetic constants associated to the catalysis between consecutive templates,
$d = 4$ is the maximum number of templates that coexist without entering the dangerous oscillatory regime. Spatial implementations of hypercycles based on cellular automaton are of little aid on this matter as it was shown (Scheuring et al., 2003) that the spatial patterns (spirals) that would supposedly increase the robustness of hypercycles against parasites (Boeljist and Hogeweg, 1991) are unstable against mutations that result in differential death rates of replicators in the cells. Monte-Carlo simulations aiming at investigating the effect of a finite number of vesicles corroborate our main conclusions.

We have also shown here from first principles that redundancy was costly in a package model: it increased the mutational load. This may not sound surprising at all from basic population genetics: we already know that other things being equal diploids end up being worse off than haploids (Crow and Kimura, 1965). What is significant is that the benefits of redundancy to lessen the assortment load are quickly overcome by the burden of deleterious mutations (Fig. 7). This suggests that selection for linkage in package models to increase the proportion of daughter cells containing a complete set of genes (see Maynard Smith and Szathmáry, 1993, 1995; Santos, 1998) was already quite strong from the very beginning.

A word is in order about the effects of the model assumptions on the main conclusions of this contribution. The instantaneous replacement of vesicles during the recolonization stage is actually a consequence of the assumption of non-overlapping generations for the vesicle dynamics (the same holds for the template dynamics as well)—a hypothesis widely used in population genetics in both deterministic and stochastic formulations. So we should not expect problems here, provided we avoid singular situations in which the entire (finite or infinite) population goes extinct in a single time step. This global extinction yields a singularity because we have assumed that the template population as well as the vesicle population are of fixed sizes, $N$ and $M \rightarrow \infty$, respectively. However, by setting $g \neq 1$ and $\tau \neq 1$ we can safely ignore this difficulty. The assumption of equal generation times for templates and vesicles—the life cycle comprehends the production of a new generation of both templates and vesicles—certainly enhances the power of group selection to maintain template coexistence. As far as coexistence is regarded, the optimal situation seems to be just the opposite one, i.e. vesicle generation times should be shorter than template generation times, but this option is out of question when the increase in the number of templates must keep pace with the increase of the vesicle, otherwise the former would be diluted out during protocell division. In any event, this new hindrance to template coexistence can be compensated for by increasing the group selection pressure. Moreover, since a discrepancy in generation times hinders coexistence and so reduces the chances of survival of the vesicle (Silvestre and Fontanari, 2005), one may expect that only templates characterized by long generation times will thrive. This is in fact the advantage of working with package models—in principle we can assume that all parameters are optimized so as to maximize the survival ability of the vesicles. According to this reasoning, however, it would be difficult to explain the selection for high values of the parameter $g$, since from the vesicle's perspective what matters is to maximize the baseline survival probability $1 - g$, i.e. minimize $g$, regardless of the presence of templates within it. This is so because Eq. (1) unnecessarily couples the baseline survival probability with the benefit conferred to the vesicle by the presence of the templates. The particular form of that equation was very convenient to our purposes, since it automatically restricts $\varkappa(k)$ to the interval $[0, 1]$, as it should be since $\varkappa$ is interpreted as a probability. A more general formulation, e.g.

$$\varkappa(k) = a + b \frac{d}{N} (k_1 k_2 \ldots k_d)^{1/d}$$

(8)

easily resolves that difficulty, since it is clear now that clonal competition will favour a high value of the parameter $b$. This formulation, however, is a bit awkward because we must be careful to not violate the constraint $\varkappa(k) < 1$ while choosing $a$ and $b$. So there is no fundamental difficulty to modify our model in order to explain selection for high values of the group selection pressure. In summary, provided the number of vesicles is sufficiently large (or small but fixed as in our Monte-Carlo simulations), our main conclusions will remain valid no matter which of our model assumptions we relax. This is the kind of robustness we see in population genetics models and there is no reason this should be different in our model. However, if the number of vesicles is small and allowed to change, so it can decrease and eventually reach zero, then the outcome will probably be different. But the study of this situation requires a totally distinct approach (coupling two stochastic dynamics), which even in the context of the well-studied problem of the evolution of altruism has not been undertaken.

Compartment (vesicle, protocell) models that explicitly invoke group selection, such as the stochastic corrector model (Szathmáry and Demeter, 1987), are sometimes advanced and motivated as alternatives to the hypercycle model. We argue, however, that group selection is an indispensable ingredient even in a more basic situation: if the subsequent stage of Eigen’s quasispecies model—compartmentalization of unlinked templates—is to be considered. In this sense, package models can be thought of as complementary to Eigen’s model in that they describe a different stage of evolution. On biological grounds the ingredient of group selection was automatically guaranteed once vesicles met the criteria for “minimal life”, i.e. a three-component system integrated by a lipid boundary, a metabolic subsystem for growing, and a genetic or informational subsystem (see, e.g. Gánti, 1987, 2003; Luisi, 1998; Szathmáry, 2003a, b). Attempts at constructing such metabolising vesicles are under way (see Szostak et al., 2001; Rasmussen et al., 2004).
Acknowledgments

This work was done during the visit of J.F.F. to Collegium Budapest in November 2003. J.F.F. is supported by CNPq and FAPESP, project 99/09644-9. M.S. is supported by a fellowship (PR2003-0431) from the Programa Nacional de Ayudas para la Movilidad de Profesores de Universidad, Ministerio de Educación, Cultura y Deporte (Spain); and by Fundación Ramón Areces (Spain). M.S. and E.S. greatly acknowledge the sponsorship from COST-European Cooperation in the field of Scientific and Technical Research, Internal Project number D27/0003/02 “Emergence and Selection of Networks of Catalytic Species” (URL: http://cost.cordis.lu/src/extranet/publish/D27WGP/d27-0003-02.cfm).

References

Scheuring, I., 2000. Avoiding Catch-22 of early evolution by stepwise increase in copying fidelity. Selection 1, 135–145.