J Mol Evol (2009) 69:568–578
DOI 10.1007/s00239-009-9281-y

Mutations and Lethality in Simulated Prebiotic Networks

Aron Inger · Ariel Solomon · Barak Shenhav · Tsviya Olender · Doron Lancet

Received: 24 August 2009 / Accepted: 8 September 2009 / Published online: 29 September 2009

Abstract The Graded Autocatalysis Replication Domain (GARD) model describes an origin of life scenario which involves non-covalent compositional assemblies, made of monomeric mutually catalytic molecules. GARD constitutes an alternative to informational biopolymers as a mechanism of primordial inheritance. In the present work, we examined the effect of mutations, one of the most fundamental mechanisms for evolution, in the context of the networks of mutual interaction within GARD prebiotic assemblies. We performed a systematic analysis analogous to single and double gene deletions within GARD. While most deletions have only a small effect on both growth rate and molecular composition of the assemblies, $\sim$10% of the deletions caused lethality, or sometimes showed enhanced fitness. Analysis of 14 different network properties on 2,000 different GARD networks indicated that lethality usually takes place when the deleted node has a high molecular count, or when it is a catalyst for such node. A correlation was also found between lethality and node degree centrality, similar to what is seen in real biological networks. Addressing double knockout mutations, our results demonstrate the occurrence of both synthetic lethality and extragenic suppression within GARD networks, and convey an attempt to correlate synthetic lethality to network node-pair properties. The analyses presented help establish GARD as a workable alternative prebiotic scenario, suggesting that life may have begun with large molecular networks of low fidelity, that later underwent evolutionary compaction and fidelity augmentation.

Keywords Origin of life · Mutations · Networks · Lethality · Synthetic lethality

Introduction

It is widely accepted that the appearance of self-reproducing protocells must have involved the emergence of networks of mutually interacting molecules, so as to resemble present-day cells (Bagley and Farmer 1991; Farmer et al. 1986; Jain and Krishna 2001; Kaneko 2002; Kauffman 1993; Shenhav et al. 2005b). What is still undecided is the nature of the specific molecules that involved in such a process. One major scenario claims that life began with a single self-replicating molecule, e.g., RNA (Gesteland et al. 2000; Gilbert 1986; Joyce 2002), which later evolved into networks instructed by the replicating biopolymers. An alternative scenario claims that as early as replicating entities have emerged, they must have constituted relatively complex molecular networks, arising via spontaneous accretion of weakly bound assemblies of simpler organic molecules (Bachmann et al. 1992; Barandiaran and Ruiz-Mirazo 2008; Benko et al. 2003; Dyson 1999; Jain and Krishna 2002; Kaneko 2002; Kauffman 1993; Luisi et al. 1999; Ruiz-Mirazo and Moreno 2004; Segre et al. 2000, 2001; Segre and Lancet 2000; Shapiro 2006; Shenhav et al. 2003; Stadler 1991; Morowitz 1992). In this alternative scenario it is further proposed that assembly reproduction directly stems from certain network

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9281-y) contains supplementary material, which is available to authorized users.

A. Inger · A. Solomon · B. Shenhav · T. Olender · D. Lancet (✉)
Department of Molecular Genetics and the Crown Human Genome Center, Weizmann Institute of Science, Rehovot 76100, Israel
e-mail: doron.lancet@weizmann.ac.il

attributes. To provide support for the "network-first" scenario one must better understand the network properties of the implicated molecular assemblies and compare them to those of contemporary living cells. We have termed this approach "early systems biology" or "systems prebiology" (Shenhav et al. 2005b) and along with others have suggested that it has potential for merging the two seemingly conflicting scenarios for prebiotic evolution (Luisi 2004; Shenhav et al. 2003).

For more than a decade we have explored the Graded Autocatalysis Replication Domain (GARD) model as a formalism for describing and simulating the behavior of early systems with mutually catalytic interaction networks (Segre et al. 1998a, 2000; Shenhav et al. 2003, 2004). In its most explored version, GARD assumes micelle-like assemblies composed of amphiphilic (lipid-like) molecules, held together by hydrophobic forces. Analogous studies directed at simulating amphiphile assemblies in the context of prebiotic evolution have been amply described (Ruiz-Mirazo and Mavelli 2007, 2008). The GARD formalism defines the probability for mutual catalytic potencies based on the behavior of combinatorial chemistry (Lancet et al. 1993, 1994; Rosenwald et al. 2002), specifying a beta matrix of non-zero rate enhancement parameters for all pairs of molecules, and defining a fully connected weighted network of interactions.

The GARD model provides a detailed dynamic description of time-dependent count changes for N_G different molecular types within the assembly. Along the time axis, a dynamic process ensues within a molecular assembly of finite size, whereby certain molecule types are enriched and others are catalytically selected against. Quasi-stationary states of the system, with idiosyncratic composition are termed composomes (Segre et al. 2000), and clusters of such in compositional space are referred to as compotypes (Naveh et al. 2004; Shenhav et al. 2007), similes of early biological species. A specific instance of GARD, defined by a particular set of mutual catalytic values within the beta matrix, may have several different compotypes, and each of these compotypes defines a weighted network of catalytic interactions (Fig. 1a).

In the basic GARD formalism, the only chemical reactions being modeled are catalyzed exchange reactions—accretion (joining) and dissociation (leaving) of molecules. The resulting dynamics involve compositional transitions that may be regarded as transient mutations (Segre and Lancet 2000; Segre et al. 1998b). When viewed within a limited time frame, the dynamics of this simple GARD model manifest graded transitions between different molecular networks, resembling an evolutionary process. In the presented work, the effect of permanent deleterious mutations in GARD networks is examined in detail, and conclusions are drawn with respect to GARD network properties that affect lethality and synthetic lethality. Some parallels are also drawn to present-day biological networks, accentuating the potential relevance of the GARD model to present-day cells, hence strengthening the case of the relevance of GARD to prebiotic scenarios.

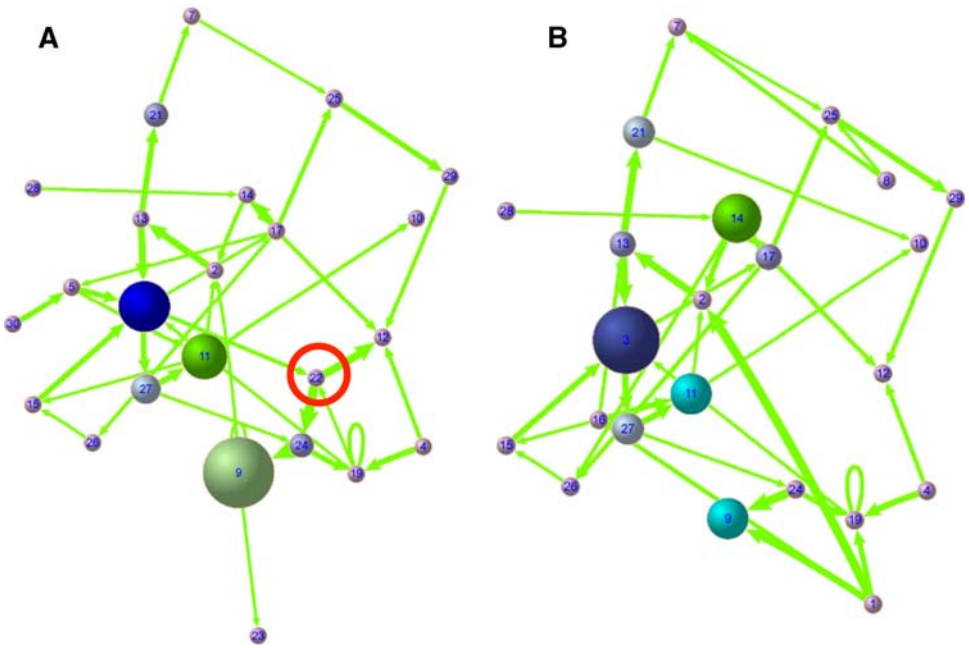

Fig. 1 A GARD network. **a** A typical wide-type compotype of the GARD system with a molecular repertoire of $N_G = 30$. Nodes sizes are scaled to molecular counts and the edges are the effective catalytic rates among them. Only molecules with a node size characterized by normalized count $NS_v > 0.3$ are shown. **b** The same GARD network, after knockout of molecule type 22 (*red circle* in **a**). An example of a network cascading effect is seen, whereby knockout of node 22 results in a decrease in the size of node 24 which it catalyzes, which subsequently leads to a decrease in the count of node 9 catalyzed by node 24

Methods

GARD Traces

Each generation in the simulation constitutes a series of molecular join–leave events governed by the set of GARD differential equations (Segre et al. 2000; Shenhav et al. 2007), which occur until a predetermined assembly size is reached, whereby random fission takes place. A simulation trace consists of a time series of pre-fission assembly compositions which define a trajectory in N_G-dimensional compositional vector space. Elements of the trace that are not within a predefined similarity distance $H > 0.9$ to the previous and next generations are considered to be drift (and not within a composome) and are removed from the trace. *K*-means clustering is then performed on the molecular compositions over all non-drift generations, resulting in a number compotypes, characterized by a lifetime represented by the number of generations it appears in the trace (Shenhav et al. 2007).

We performed GARD simulations using the procedures described in Segre et al. (2000). The main simulation engine is a Matlab program TGS_AGARD, based on the Gillespie algorithm (Gillespie 2001), and runs on both windows and Linux platforms. For each simulation we used a different beta matrix randomly selected from a log normal distribution with $\mu = -4$ and $\sigma = 4$. The parameters for the GARD runs are as follows; k_f (*k* forward) $= 1 \times 10^{-1}$, k_b (*k* backward) $= 1 \times 10^{-4}$, N_G (molecular repertoire size) $= 30$ or 100, number of generations $= 500$ or 1,000, respectively, N_0 (split-size) $= 45$ or 150, respectively, range of number of clusters tested for *k*-means 1–10.

Network Node Properties

We transformed the GARD trace (Shenhav et al. 2005a), constituting 500 growth-split generations into a network by performing the following procedure. First we defined the effective matrix E as follows. Let $E_{ij} = \text{Beta}_{ij} \times C_{W1i}$, where beta is the $N_G \times N_G$ matrix defining the catalytic strength for all interactions between the N_G species, and C_{W1} is the vector describing the counts of all N_G species in the most frequent wild-type compotype (this is a specific instance of the GARD vector of molecular counts generally termed n_i). Then we defined the adjacency matrix A where $A_{ij} = 1$ when $E_{ij} > \text{mean}(E) + \text{STD}(E)$ and 0 otherwise. The adjacency matrix defines a directed graph with up to N_G nodes, and binary edges. We then calculated the following node properties on this graph for each node. Node pair properties are taken as the mean of the single node properties, except three properties pertaining to node pairs (end of Table 1S).

Definition of Lethality and Synthetic Lethality

Four fitness measures were defined as detailed in Table 2S. A node is considered to be lethal if its H fitness measure is less than 0.8. The graded synthetic lethality of nodes i and j is calculated as the synthetic fitness $H_s = H_{ij}/H_iH_j$. A pair of nodes i and j are considered to be synthetically lethal if they satisfy the following three criteria: (1) $H_i > 0.8$, $H_j > 0.8$, (2) $H_{ij} < 0.8$, and (3) $H_{ij} < 0.95 \times H_iH_j$. These equations are based on standard formalism of genetic epistasis (e.g., DeLuna et al. 2008). The above definitions are only useful for fitness measures for which the double knockout fitness is equal to the product of the single knockout fitnesses when there is no synthetic interaction. Of the four fitness measures described previously, only H satisfies this condition.

Results

Networks in the GARD Model

GARD is a chemical kinetics model for simulating the behavior of molecular mutually catalytic sets. The GARD model has been shown to be equivalent to a network of interactions among molecules within a non-covalent assembly (e.g., a micelle) (Segre et al. 2000; Shenhav et al. 2005b) (Fig. 1a). It may be regarded as an emulation of a gene network, whereby the global GARD fate may be thought of as emulating a phenotype (Shenhav et al. 2005b). By altering the underlying interaction matrix of the model, we are able to simulate genetic knockouts and scrutinize their emulated phenotypic effects. Although GARD represents a very simple network of mutual catalysis, our hypothesis is that it might provide valuable quantitative insights for understanding knockouts in the interaction networks of real cells.

A specific run of the GARD simulation is defined by a beta matrix of kinetic parameters, randomly generated from a log-normal distribution (Lancet et al. 1993, 1994; Rosenwald et al. 2002), which defines the catalytic interactions among the different molecules (see Methods). By assigning a cut-off threshold, the beta matrix is transformed into an adjacency matrix which defines a directed graph network. The nodes in the network are molecules and the edges represent the effective mutually catalytic interactions.

To assess some of the properties of the wild-type GARD networks, we simulated GARD traces with 500 generations each, for 2,000 different GARD networks with molecular repertoire sizes of $N_G = 30$. For comparison we also ran 1,000 networks with $N_G = 100$. Interestingly, we saw the appearance of a power law for the molecular counts

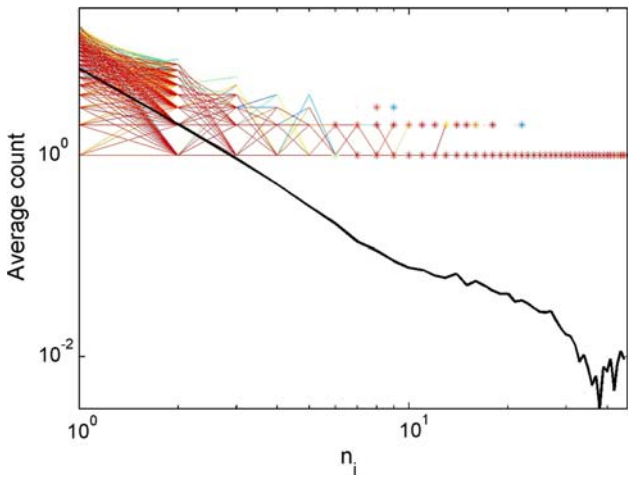

Fig. 2 Power law distribution of molecular counts for $N_G = 30$. The molecular counts C_{Wj} (Table 2S) within the compotypes that were discovered in the 2000 runs of GARD. Colors indicate different compotypes, with the highest molecular counts indicated by an *asterisk*. Average count is 1.5. The *black line*, representing the average frequency of the different molecular counts over all compotype, shows a high ($R = -0.98$) log–log linear fit obeying a power-law $P(X) \sim X^{-1.8}$

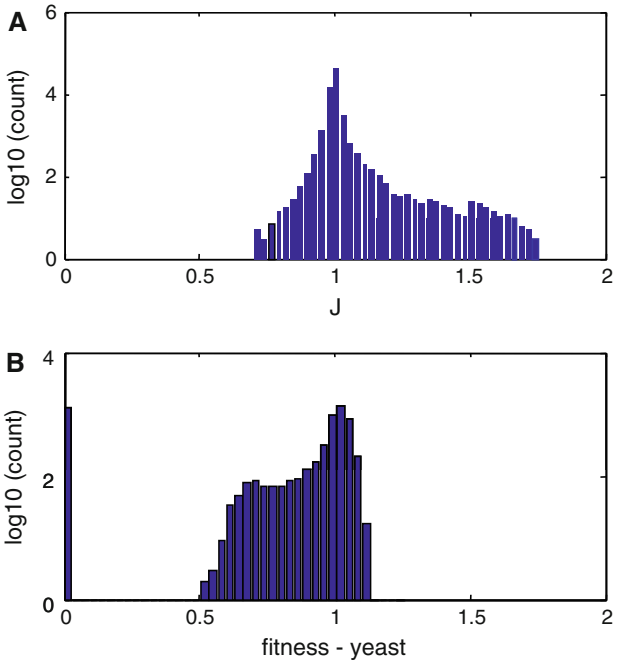

Fig. 3 A comparison of fitness distributions. **a** Mutated GARD, where the values shown are the J fitness measure (Table 2S), representing the change of GARD assembly join/leave rates following a knockout. **b** Yeast fitness values (Steinmetz et al. 2002) for 5,910 gene single knockouts. The fitness measure shown is the change in growth rate following gene deletion

(Figs. 2, 5S), a phenomenon akin to the distribution of concentrations of m-RNA transcripts in biological cells (Kaneko 2003; Lu and King 2009). These wild-type simulations served as a baseline for the forthcoming mutation analysis.

GARD Knockouts, Fitness, and Lethality

Knockout-like events are introduced into the GARD network by setting to zero the entire row and column of a given molecule in the beta matrix. This has the effect of eliminating all catalytic interactions with the mutated molecule. In the main analysis, we built upon the wild-type GARD simulations with $N_G = 30$, as described above, and performed for all possible single knockouts of the individual molecular nodes. As in the case of cellular networks in biological organisms, knockout events may influence the network in non-trivial, often unpredictable ways based on alternative pathways and cascading effects (Fig. 1b).

We then examined the effect of a knockout on the overall behavior of the simulated GARD network, employing measures for emulated fitness and lethality. In real cells, such as the yeast *Saccharomyces cerevisiae*, the fitness of a mutated system can be monitored by parameters such as changes of the growth rate, colony size or cell morphology, with respect to the wild-type. In the GARD model, the relevant phenotypic change was defined by comparing the compotypes between wild-type and knockout. We used four different graded measures for fitness (Methods and Table 2S), as indicated in Fig. 2S, as well as a binary definition of lethality based on a predefined threshold value for the fitness (see Methods), as described below.

Using the fitness measure H, related to the similarity of compotypes before and after knockout, we defined lethality using a threshold value $H < 0.8$. This threshold was selected based on the leveling off in the logarithmic distribution of H around this value (Fig. 1S-A). Unlike for yeast, where there is a clear distinction between lethal and deleterious non-lethal mutations (Fig. 3b), in the GARD network no such obvious demarcation was seen (Figs. 3a, 1S). Figure 4a shows the propensity of lethal knockouts in the 2,000 simulation experiments performed. It may be seen that the most probable value is 2, implying that typically most networks contained only two essential nodes, and only few showed lethality for up to 8 of the examined 30 knockouts per network.

The probability distribution for different fitness measures for GARD knockouts is shown in Figs. 3 and 1S. Focusing on the example of Fig. 3a, it is seen that the fitness measure J, which represents the change of GARD assembly growth rates following knockout, has a peak at $J = 1$ (no change), demonstrating the robustness of GARD networks to knockouts. Still, a substantial propensity exists for both $J < 1$ and $J > 1$, respectively representing diminished growth ("sickness") and enhanced growth rate following knockout. Such a bilateral distribution is also

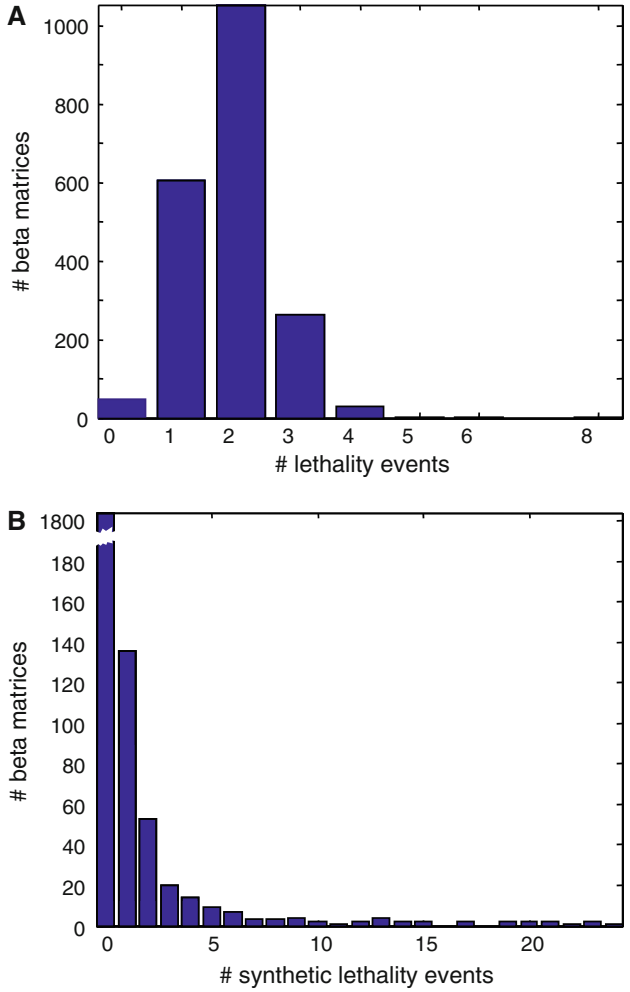

Fig. 4 Propensity of lethality and synthetic lethality in GARD networks. **a** A distribution of the number of lethal nodes per network with $N_G = 30$. **b** A distribution of the number of synthetic lethality events per network with $N_G = 30$. Data are as described in the "Methods" and the "Results", and in Fig. 6

observed for laboratory measurements of growth rate in yeast (Fig. 3b). This indicates a similarity between simulated GARD networks and real world cellular networks, whereby while most knockouts have negative effect, a certain fraction may lead to enhanced fitness (de Visser et al. 2003; Giaever et al. 2002). A similar bilaterality is observed for another two fitness measures (Fig. 1S-B, C). Note that the compotype similarity fitness measure (Fig. 1S-A) obeys, by definition, $H \leq 1$.

GARD Fitness and Network Properties

Our results clearly show the occurrence of lethality, sickness, and enhancement due to knockouts in the GARD network. To further study the underlying mechanism which may govern the sensitivity of GARD assembly dynamics to a knockout, we calculated 14 different network properties for each network and node. We evaluated the correlation between the various properties of the node and the fitness score for the mutation of that node (Fig. 2S). We then examined some of the individual network properties in greater detail. The network property most highly correlated with fitness is the size of the knockout node, i.e., the count of the molecule in the wild-type. The high correlation between the node size and fitness is observed for all fitness measures, but is particularly pronounced for the compotype similarity H (corr = 0.78, Fig. 2S).

Figure 5 shows a scatter plot of the size of the knockout node and the resulting fitness of the mutant for two different fitness measures, H and H'. For the H' fitness measure we see that mutations of large node sizes always lead to a drastic reduction in fitness whereas mutations of small nodes sometimes reduce the fitness and may sometimes lead to enhancement, although they usually have only a small effect on the fitness. For the H fitness measure we also see that when the knockout node size is large, the mutant fitness is low without exception.

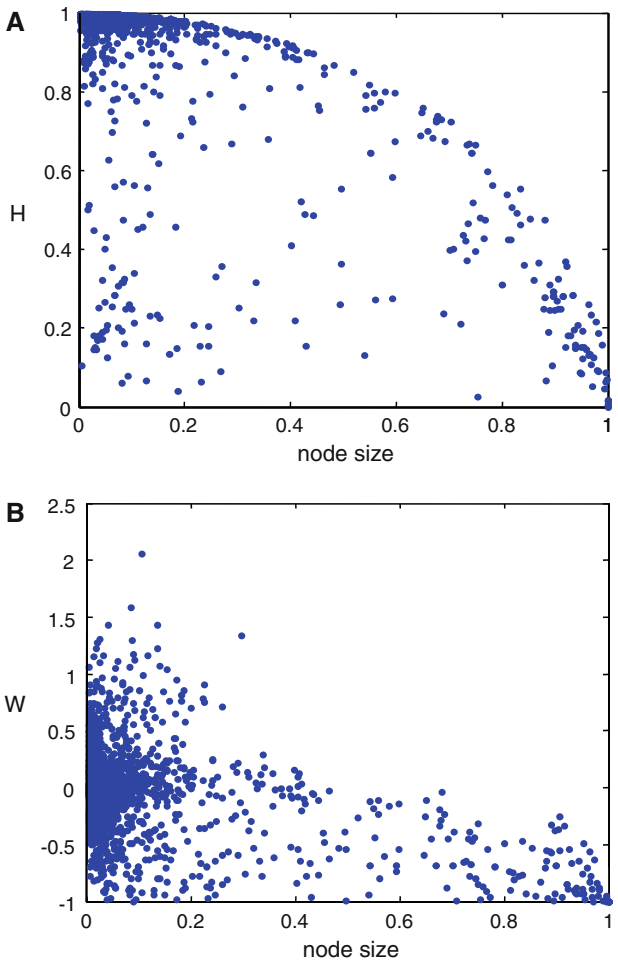

Fig. 5 Correlation between node size and fitness. **a** For the H fitness measure. **b** For the H' fitness measure. Data as in Fig. 6

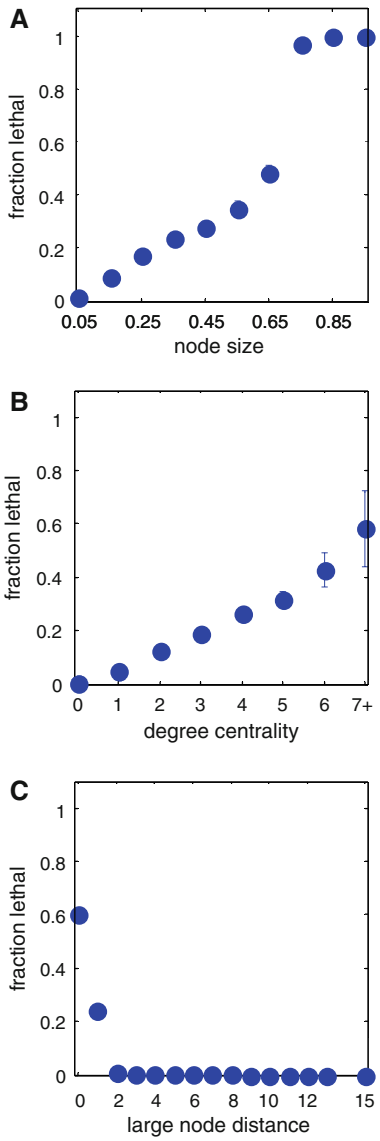

Fig. 6 Correlation between selected network properties and lethality. The calculation was performed using 2,000 different GARD networks of repertoire size $N_G = 30$ (total 60,000 mutants). Network properties are **a** node size, **b** degree centrality, and **c** distance from a large node. See Table 1S for definition of the network properties

To further explore the effects of mutations, we made use of the 3-criteria definition of lethality (Methods). Plotting the fraction of lethal cases vs. node size shows a significant correlation (Fig. 6a), including the occurrence of nearly 100% lethality above a value of node size of 0.8.

An intuitive rationalization for the observation that removing a molecule with a large count in the wild-type generates a very dissimilar mutant may be brought up for the *H* fitness measure: in this case the compositional vector is appreciably modified, so that the scalar product with the wild-type vector is bound to be low. Rewardingly, a similar correlation (albeit weak) has been reported between gene expression levels and the deleterious effect on viability in the yeast *Saccharomyces cerevisiae* (Pal et al. 2003).

Node size dependence may be one of the underlying explanatory mechanisms for mutation sensitivity. However, there are also many points in Fig. 5 with a low node size that also yield a low fitness mutation. A rationalization may come from the data in Fig. 6c, showing the lethal fraction for nodes based on their graph distance from any large node. A distance of zero means the node in question is itself a large node, already shown to be highly lethal when mutated. A distance of one means the node itself is not large, but it catalyzes the entry of a molecular node which is large in the wild-type. Thus, such "enabling" nodes also appear to have a tendency to be lethal when mutated. This is not counter-intuitive, as by removing these nodes, the dominating molecules in the wild-type are not sufficiently catalyzed. In fact, for the GARD simulation, the two conditions of being either a large node or a catalyst for a large node are responsible for almost all of the observed lethality, as only negligible lethality is seen for nodes with distance two or larger (Fig. 6c). The capacity to perform such analysis is indicative of an advantage of GARD networks, in which nodes have a size attribute. This is in contrast to a significant fraction of biological network analyses in which node size is not included (Platzer et al. 2007).

We further performed a correlation analysis between the lethal fraction and degree centrality for GARD networks (Fig. 6b). A well-defined linear relationship is seen, whereby a higher degree for the knocked-out results in more lethality. A similar relationship between node lethality and degree centrality was previously reported in certain real biological networks (Barabasi and Oltvai 2004; Rodrigues and Costa Lda 2009), although claims to the contrary have also been voiced (He and Zhang 2006; Terry 1992; Zotenko et al. 2008).

To examine the role of network size on our results, we simulated an additional 1,000 GARD networks of the size $N_G = 100$ and repeated the same analyses (Fig. 2S-B). This yielded the same observation, i.e., that lethality is correlated with a high node size, short distance from large node and degree centrality (Fig. 3S), although the overall correlation scores were lower.

Synthetic Lethality in GARD

We asked whether GARD networks also manifest the highly significant phenomenon of synthetic lethality. In the simplest definition, two genes are considered synthetically lethal if knocking out either of them has no phenotype while inactivating both leads to death (Hartman et al. 2001). Such synergy is, however, often recorded on a more graded scale, whereby in synthetic lethality

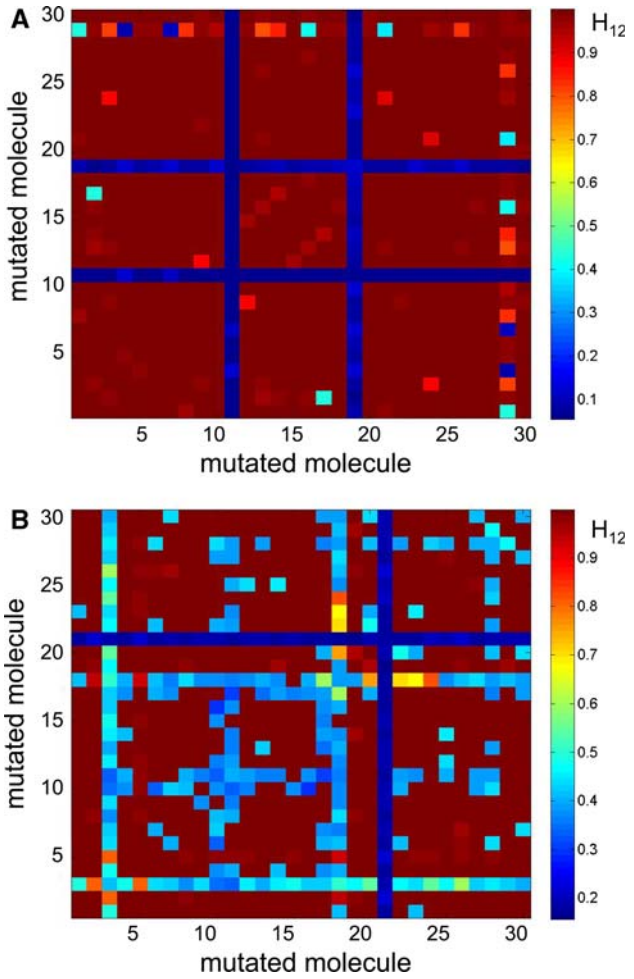

Fig. 7 Synthetic lethality within the GARD networks. The figure shows two specific fitness matrices, obtained from all possible single and double knockouts in two specific GARD networks of a repertoire size $N_G = 30$. Fitness measures of single mutations are shown along the diagonal, and off-diagonals are fitness measures of double knockouts. The fitness was calculated using the H measure (color bar) where a value of 1 represents no compotype change

the fitness of the double mutant is significantly lower than the product of the individual fitnesses. We performed all 435 possible double knockouts of pairs of molecules in the 2,000 different GARD networks with $N_G = 30$. Using this database, we first recorded the values of H_{ij} for each double knockout i and j, a parameter that indicates its compotype similarity to the wild-type, in analogy to H for single mutations.

Figure 7 shows typical results for the double mutation matrix: dark or light blue rows/columns show cases of single lethality, while isolated off-diagonal blue pixels are indicative of synthetic lethality. It is evident that GARD networks can show such node interaction, as is the case for cellular networks (e.g., in yeast, see Boone et al. 2007). Interestingly, Fig. 7b also shows cases of extragenic suppression, whereby for a lethal mutation (such as molecular

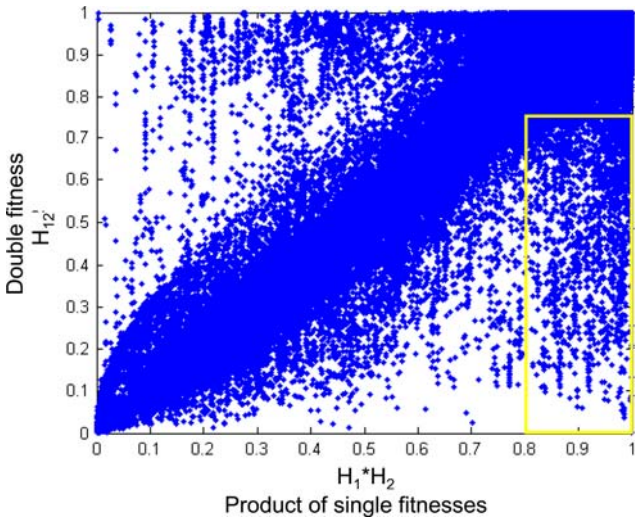

Fig. 8 Double knockout fitness. The synthetic fitness H_{ij} is plotted against the product of the corresponding single fitness values ($H_i \cdot H_j$) for all possible pairs in 2,000 different GARD networks of repertoire size $N_G = 30$. Points which fall along the diagonal represent cases with minor synthetic interaction. Knockout pairs which are certain to be synthetic lethal as defined by the 3-criteria formula (Methods) are within the delimited lower right triangle, and synthetic enhancement are symmetrically disposed on the upper left

component 3) the double mutant is viable (in this example—via a second mutation in components 2 or 5).

We further provide a measure of synthetic lethality, using a definition based on a cutoff of $H \leq 0.8$ (see Methods). Using this definition, we analyzed the propensity of synthetic lethality in GARD networks (Fig. 4b). It is evident that about 10% of all the networks examined show at least one case of synthetic lethality, and more that 20 ($\sim$1%) show >10 events per network. The overall rate of synthetic lethality in GARD networks of size $N_G = 30$ is about 0.1%, as compared to about 1.5% in yeast (Tong et al. 2004).

A comprehensive analysis of all the knockouts performed is shown in Fig. 8, where H_{ij} is plotted against the product of the two single mutant fitnesses (H_iH_j). Points near the diagonal represent cases of synthetic fitness $H_s = H_{ij}/H_iH_j \approx 1$ (no interaction). Values of H_s lower than one indicate a deleterious synthetic interaction and values greater than one indicate synthetic enhancement. Using the more specific 3-criteria definition of synthetic lethality (Methods), all points within the labeled rectangle in the right corner represent synthetic lethality by this definition.

GARD Network Properties Affecting Synthetic Lethality

We further searched for the conditions in which synthetic lethality occurs. For that purpose we calculated 17 node

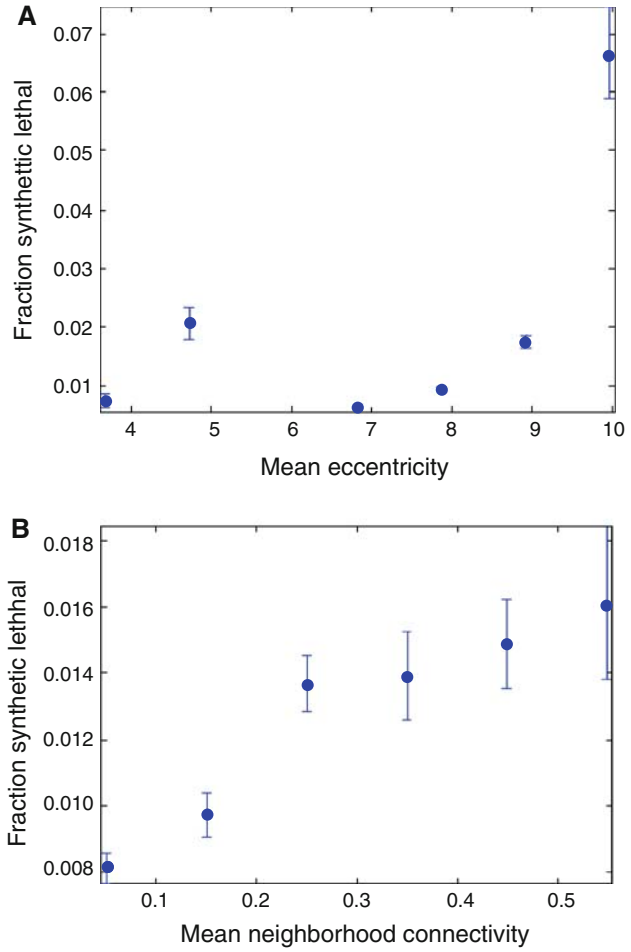

Fig. 9 Fraction of synthetic lethality cases in GARD networks for different values of **a** mean eccentricity and **b** mean neighborhood connectivity (See Table 1S)

pair network properties, in an analogous fashion to what was done for individual nodes and lethality. We then examined the correlation between the network node pair properties and synthetic lethality. The results for some of these network properties are shown in Figs. 9 and 4S. However, unlike for lethality, the correlations which we detected in general had weak statistical significance. A noticeable 7-fold enhancement in the propensity of synthetic lethality is seen at the higher mean eccentricity values (Fig. 9a), representing nodes that are far from the center of the network. In parallel, a rather linear 2-fold increase in synthetic lethality is observed as a function of mean neighborhood connectivity, indicating that nodes belonging to well-connected neighborhoods have a slightly increased propensity for synthetic lethality. Further, we observed that nodes that have highly similar outgoing (but not incoming) catalytic partners are more likely to be synthetically lethal (Fig. 4S). This hints at the possibility of a redundancy mechanism in action.

Discussion

Despite considerable scrutiny and dispute, the mechanism of life's origin remain largely undeciphered. A great challenge is bridging from models for the very early conjectural steps in life's emergence toward minimal extant life, e.g., a small bacterium. Some hypothetical scenarios bring forth the notion that the earliest living entities shared similar chemical constituents with present-day life. This approach is exemplified by RNA-first (Gesteland et al. 1999; Gilbert 1986; Orgel 1998) and protein-first scenarios (Fox 1991; Lee et al. 1996), whereby biopolymers similar to those found in cellular life today were generated abiotically, and served as functional building blocks for information transfer and catalysis in early protocells. In contrast, other approaches (Dyson 1982; Etxeberria and Ruiz-Mirazo 2009; Jain and Krishna 2001, 2002; Kauffman 1993; Ruiz-Mirazo and Mavelli 2007; Segre et al. 2000, 2002; Shapiro 2006) relax the constraints on the molecular make-up of early living entities, implicating diverse protoliving chemistries. In these latter scenarios attempts are made to provide evidence for similarity to present life not at the level of shared specific chemical repertoires, but at the level of shared operating principles.

The GARD model analyzed here belongs to the second class of models. Only very broad chemical constraints are invoked on the constituents of GARD assemblies: they are assumed to be amphiphiles of any kind, as long as they are present in the environment, and spontaneously generate micelle-like structures. An important operating principle previously shown to be shared between GARD and present-day life is a capacity to undergo reproduction-like processes. This general concept is embodied in several additional approaches indicating information transfer in entities that do not explicitly contain templating biopolymers (Etxeberria and Ruiz-Mirazo 2009; Jain and Krishna 2001, 2002; Kaneko 2002, 2003; Ruiz-Mirazo and Moreno 2004) and display autonomy and self construction. In GARD, it is manifested in the transmission of compositional information from an ancestor assembly to its progeny. We also pointed out that spontaneous compositional changes may be viewed as mimicking mutations. However, these mutational events are fleeting, as subsequent dynamic changes may bring back a molecule type which has been temporarily lost. In the present article, we explore a stable form of mutations, resembling gene deletion (knockout), changes involving the permanent disappearance of molecular types. In terms of planetary reality, such changes may be viewed as stemming from changes in the environmental molecular repertoire (Shenhav et al. 2007), or from a transition of an assembly from one microenvironment to another (Monnard et al. 2003). Yet, in network formalism, they are equivalent to node deletion, akin to gene knockout.

A systematic analysis of GARD's response to such deletions provides two types of information. On the one hand, it helps one be further convinced that GARD shows similarity of dynamic behavior to present life. This includes the bilateral distribution of fitness values upon single knockouts, the dependence of lethality on degree centrality and on node size, as well as the relation between the internode similarity of outgoing catalysis and synthetic lethality. A similar approach was previously used (Jain and Krishna 2002) to analyze the emergence of extinction events in a model involving a network of molecular entities. The results likewise indicated that "crashes" of the modeled network occur upon the extinction of a keystone molecule.

On the other hand, an analysis of the relationship between GARD network properties and the response to mutations may afford a better understanding of extant biological networks. The great amount of systems biology scrutiny that has been performed in the last decade on the characteristics of biological networks (Alm and Arkin 2003; Barabasi and Oltvai 2004; Monk 2003; Newman 2003; You 2004) involved unweighted (binary or Boolean) networks, in which a specific node is either connected or not to some others. Widely used network attributes such as degree distribution, mean path length, and clustering coefficient, are based on counting these all-or-nothing interaction edges.

However, there are many examples of real biological networks in which nodes are connected in a graded or weighted fashion, e.g., when continuous measures such as affinity or catalytic enhancement are at work. A full representation of such gradations is often hindered by the lack of information, but also because of the sheer complexity of realistic biological networks. This is exemplified by protein–protein interaction networks (Platzer et al. 2007) that are largely treated as binary. In GARD networks, on the other hand, nodes are weighted (representing molecular counts) and edges are directed and weighted (catalysis is considered in either direction and is associated with an intensity value). GARD, as a toy model, is sufficiently simple to allow comprehensive analysis of all mutations in thousands of instances of such graded networks. As a result, GARD simulations may provide a rich set of statistics which can be used to examine the effects of knockouts in ways that are not possible in yeast or other cells. Insights gleaned from GARD may become helpful when more information is available for real cellular networks, enabling analyses with weighted nodes and edges.

We have recently applied these concepts to a specific example with biological significance. This relates to the SYNLET project consortium (http://synlet.izbi.uni-leipzig.de/), which attempts to identify synthetically lethal partners for genes involved in tumorigenesis and tumor drug resistance as means to devise therapy (Iglehart and Silver 2009; Kaelin 2005; Kennedy and D'Andrea 2006). GARD results helped elucidate some of the properties of real cellular networks, and their relationships to synthetic lethality.

Inferences from GARD should, however, be treated with caution, as in some key respects GARD networks are different from those of present-day cells. For example, while biological networks are scale-free, their degree distribution following a power law (Barabasi and Oltvai 2004; Jeong et al. 2000; Siegal et al. 2007), GARD networks tend to behave more like random networks, typified by a most probable node degree value (Shenhav et al. 2005b). On the other hand, GARD network node sizes (molecular counts in compotypes) do show a power law distribution, as we show here. A power law distribution, also known as Zipf law, was suggested to describe the relative abundance of within m-RNAs repertoires in living cells (Kaneko 2003). A later analysis suggests that a power-law, or a related distribution, serves as a guiding rule also for the amounts within other molecular repertoires such as proteins or metabolites (Lu and King 2009), which are even more analogous to GARD molecular types.

In present-day cellular networks, connectivity is sparse due to the high recognition specificity of the components attained along evolution. Thus, in a protein interaction network with many thousands of nodes, most nodes have no more than a few dozen interaction partners, mutually recognized at high affinity and specificity. In contrast, at life's origin it stands to reason that recognition was much more promiscuous. In particular, in prebiotic scenarios as proposed (Dyson 1999; Kauffman 1993; Morowitz 1992; Segre et al. 2001; Segre and Lancet 2000; Shapiro 2006; Wachtershauser 1990), prior to the appearance of biopolymers such as globular proteins and long RNAs, simpler organic molecules may have played key roles in mutual recognition and catalytic interactions. In such circumstances, molecular constituents could have much larger numbers of interacting partners, and the corresponding networks would encompass very high average degree values. Furthermore, most interacting pairs would have affinities in the range what is presently considered non-specific binding.

The GARD scenario for life's origin therefore involves evolutionary transitions from large, low-specificity networks to more compact, high-specificity, and high fidelity ones. This is in contrast to the RNA-first scenario that begins with one or very few molecule types with relatively high-interaction fidelity, such that affords replication via base-paring (Fig. 10). Subsequently recruitment of further components would lead to more life-like mutual interaction networks. Future studies will have to judge which of these dichotomous scenarios is more relevant to the way by which life first appeared, and whether a synthesis between the two is feasible.

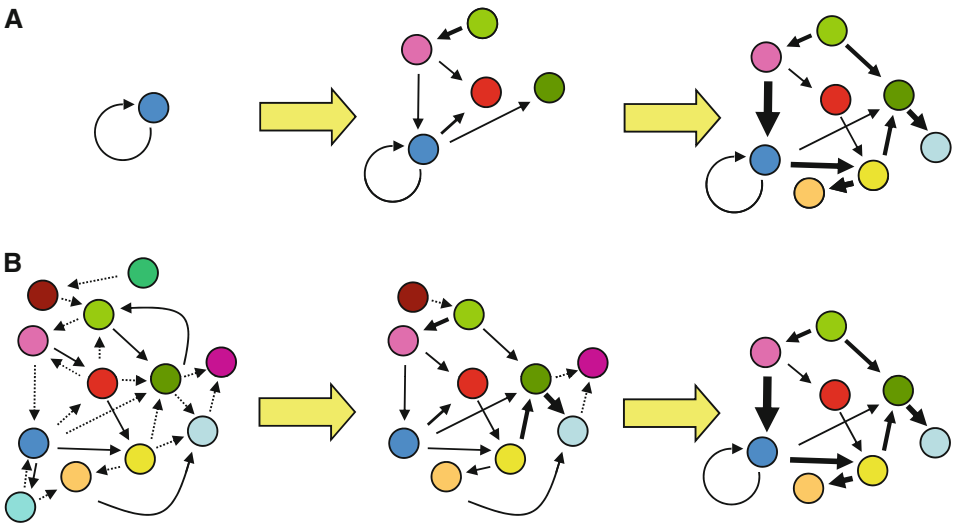

Fig. 10 Two alternative views of the progression from early molecular entities to more life-like networks. **a** Early evolutionary steps involve one molecule type with template replication capacity, progressing gradually toward the evolutionary target by increasing network size. **b** Early networks with large node count and low interaction fidelity progress gradually by node weeding and fidelity enhancement toward the same evolutionary target

Acknowledgments This work is supported by the EU Specific Targeted Research Project consortium "Regulatory Control Networks Synthetic Lethality" (SYNLET, Grant 043312) and by the Crown Human Genome Center at the Weizmann Institute of Science. The authors wish to thank to Y. Pilpel, N. Barkai, and I. Tirosh for the useful discussions.

References

Alm E, Arkin AP (2003) Biological networks. Curr Opin Struct Biol 13:193–202

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

Bagley RJ, Farmer DJ (1991) Spontaneous emergence of a metabolism. In: Langton CG, Taylor C, Farmer JD, Rasmussen S (eds) Artificial life II. Addison-Wesley, Redwood City, pp 93–140

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

Barandiaran X, Ruiz-Mirazo K (2008) Modelling autonomy: simulating the essence of life and cognition. Introduction. Biosystems 91:295–304

Benko G, Flamm C, Stadler PF (2003) A graph-based toy model of chemistry. J Chem Inf Comput Sci 43:1085–1093

Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. Nat Rev Genet 8:437–449

de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, Gibson G, Hansen TF, Krakauer D, Lewontin RC, Ofria C, Rice SH, von Dassow G, Wagner A, Whitlock MC (2003) Perspective: evolution and detection of genetic robustness. Evolution 57:1959–1972

DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R (2008) Exposing the fitness contribution of duplicated genes. Nat Genet 40:676–681

Dyson FJ (1982) A model for the origin of life. J Mol Evol 18:344–350

Dyson FJ (1999) Origins of life. Cambridge University Press, Cambridge, UK

Etxeberria A, Ruiz-Mirazo K (2009) The challenging biology of transients. A view from the perspective of autonomy. EMBO Rep 10(Suppl 1):S33–S36

Farmer JD, Kauffman SA, Packard NH (1986) Autocatalytic replication of polymers. Physica 22D:50–67

Fox SW (1991) Synthesis of life in the lab? Defining a protoliving system. Q Rev Biol 66:181–185

Gesteland RF, Cech TR, Atkins JF (1999) The RNA world, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Gesteland RF, Cech TR, Atkins JF (2000) The RNA world. Cold Spring Harbor Laboratory, Cold Spring Harbor

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418:387–391

Gilbert W (1986) The RNA world. Nature 319:618

Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys 115:1716–1733

Hartman JLt, Garvik B, Hartwell L (2001) Principles for the buffering of genetic variation. Science 291:1001–1004

He X, Zhang J (2006) Why do hubs tend to be essential in protein networks? PLoS Genet 2:e88

Iglehart JD, Silver DP (2009) Synthetic lethality—a new direction in cancer-drug development. N Engl J Med 361:189–191

Jain S, Krishna S (2001) A model for the emergence of cooperation, interdependence, and Structure in evolving networks. Proc Natl Acad Sci USA 98:543–547

Jain S, Krishna S (2002) Large extinctions in an evolutionary model: the role of innovation and keystone species. Proc Natl Acad Sci USA 99:2055–2060

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. Nature 407: 651–654

Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–221
Kaelin WG Jr (2005) The concept of synthetic lethality in the context of anticancer therapy. Nat Rev Cancer 5:689–698
Kaneko K (2002) Kinetic origin of heredity in a replicating system with a catalytic network. J Biol Phys 28:781–792
Kaneko K (2003) Recursiveness, switching, and fluctuations in a replicating catalytic network. Phys Rev E Stat Nonlin Soft Matter Phys 68:031909
Kauffman SA (1993) The origins of order—self-organization and selection in evolution. Oxford University Press, New York
Kennedy RD, D'Andrea AD (2006) DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. J Clin Oncol 24:3799–3808
Lancet D, Sadovsky E, Seidemann E (1993) Probability model for molecular recognition in biological receptor repertoires—significance to the olfactory system. Proc Natl Acad Sci USA 90:3715–3719
Lancet D, Kedem O, Pilpel Y (1994) Emergence of order in small autocatalytic sets maintained far from equilibrium—application of a probabilistic receptor affinity distribution (RAD) model. Berichte Der Bunsen-Gesellschaft-Phys Chem Chem Phys 98:1166–1169
Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. Nature 382:525–528
Lu C, King RD (2009) An investigation into the population abundance distribution of mRNAs, proteins, and metabolites in biological systems. Bioinformatics 25:2020–2027
Luisi PL (2004) Introduction to COST27, special issue. Orig Life Evol Biosph 34:1–2
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
Monk NA (2003) Unravelling nature's networks. Biochem Soc Trans 31:1457–1461
Monnard PA, Kanavarioti A, Deamer DW (2003) Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. J Am Chem Soc 125:13734–13740
Morowitz HJ (1992) The beginnings of cellular life. Yale University Press, New Haven
Naveh B, Sipper M, Lancet D, Shenhav B (2004) Lipidia: an artificial chemistry of self-replicating assemblies of lipid-like molecules. In: Proceeding of the 9th international conference on the simulation and synthesis of living systems (ALIFE9), Boston, Massachusetts, pp 501–506
Newman MEJ (2003) The structure and function of complex networks. Siam Rev 45:167–256
Orgel LE (1998) The origin of life—a review of facts and speculations. TIBS 23:491–495
Pal C, Papp B, Hurst LD (2003) Genomic function: rate of evolution and gene dispensability. Nature 421:496–497 discussion 497–498
Platzer A, Perco P, Lukas A, Mayer B (2007) Characterization of protein-interaction networks in tumors. BMC Bioinform 8:224
Rodrigues FA, Costa Lda F (2009) Protein lethality investigated in terms of long range dynamical interactions. Mol Biosyst 5:385–390
Rosenwald S, Kafri R, Lancet D (2002) Test of a statistical model for molecular recognition in biological repertoires. J Theor Biol 216:327–336
Ruiz-Mirazo K, Mavelli F (2007) Question 7: modelling minimal 'lipid-peptide' cells. Orig Life Evol Biosph 37:433–437
Ruiz-Mirazo K, Mavelli F (2008) On the way towards 'basic autonomous agents': stochastic simulations of minimal lipid-peptide cells. Biosystems 91:374–387
Ruiz-Mirazo K, Moreno A (2004) Basic autonomy as a fundamental step in the synthesis of life. Artif Life 10:235–259
Segre D, Lancet D (2000) Composing life. EMBO Rep 1:217–222
Segre D, Lancet D, Kedem O, Pilpel Y (1998a) Graded autocatalysis replication domain (GARD): kinetic analysis of self-replication in mutually catalytic sets. Orig Life Evol Biosph 28:501–514
Segre D, Pilpel Y, Lancet D (1998b) Mutual catalysis in sets of prebiotic organic molecules: evolution through computer simulated chemical kinetics. Physica A 249:558–564
Segre D, Ben-Eli D, Lancet D (2000) Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proc Natl Acad Sci USA 97:4112–4117
Segre D, Ben-Eli D, Deamer DW, Lancet D (2001) The lipid world. Orig Life Evol Biosph 31:119–145
Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99:15112–15117
Shapiro R (2006) Small molecule interactions were central to the origin of life. Q Rev Biol 81:105–125
Shenhav B, Segre D, Lancet D (2003) Mesobiotic emergence: molecular and ensemble complexity in early evolution. Adv Complex Syst 6:15–35
Shenhav B, Kafri R, Lancet D (2004) Graded artificial chemistry in restricted boundaries. In: Proceedings of 9th international conference on the simulation and synthesis of living systems (ALIFE9). Boston, Massachusetts, USA
Shenhav B, Bar-Even A, Kafri R, Lancet D (2005a) Polymer GARD: computer simulation of covalent bond formation in reproducing molecular assemblies. Orig Life Evol Biosph 35:111–133
Shenhav B, Solomon A, Lancet D, Kafri R (2005b) Early systems biology and prebiotic networks. Trans Comput Syst Biol LNCS 3380:14–27
Shenhav B, Oz A, Lancet D (2007) Coevolution of compositional protocells and their environment. Philos Trans Roy Soc B 362:1813–1819
Siegal ML, Promislow DE, Bergman A (2007) Functional and evolutionary inference in gene networks: does topology matter? Genetica 129:83–103
Stadler PF (1991) Dynamics of autocatalytic reaction networks. IV: inhomogeneous replicator networks. Biosystems 26:1–19
Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW (2002) Systematic screen for human disease genes in yeast. Nat Genet 31:400–404
Terry MA (1992) Writing a multiple-choice test question. J Am Osteopath Assoc 92:112–114 123
Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C (2004) Global mapping of the yeast genetic interaction network. Science 303:808–813
Wachtershauser G (1990) Evolution of the first metabolic cycles. Proc Natl Acad Sci USA 87:200–204
You L (2004) Toward computational systems biology. Cell Biochem Biophys 40:167–184
Zotenko E, Mestre J, O'Leary DP, Przytycka TM (2008) Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. PLoS Comput Biol 4:e1000140