Running title: Dimer synthesis in reproducing assemblies

Polymer GARD: computer simulation of covalent bond formation in reproducing molecular assemblies

Barak Shenhav, Arren Bar-Even, Ran Kafri and Doron Lancet*

Department of Molecular Genetics and the Crown Human Genome

Center, the Weizmann Institute of Science, Rehovot 76100, Israel

 * Corresponding author, at: doron.lancet@weizmann.ac.il

Phone: 972-8-9343683 or 972-8-9344121; Fax: 972-8-9344487

Abstract. The basic Graded Autocatalysis Replication Domain (GARD) model consists of a repertoire of small molecules, typically amphiphiles, which join and leave a non-covalent micelle-like assembly. Its replication behavior is due to occasional fission, followed by a homeostatic growth process governed by the assembly's composition. Limitations of the basic GARD model are its small finite molecular repertoire and the lack of a clear path from a 'monomer world' towards polymer-based living entities. We have now devised an extension of the model (polymer GARD or P-GARD), where a monomer-based GARD serves as a 'scaffold' for oligomer formation, as a result of internal chemical rules. We tested this concept with computer simulations of a simple case of monovalent monomers, whereby more complex molecules (dimers) are formed internally, in a manner resembling biosynthetic metabolism. We have observed events of dimer 'take-over' - the formation of compositionally stable, replication-prone quasi stationary states (composomes) that have appreciable dimer content. The appearance of novel metabolism-like networks obeys a time-dependent power law, reminiscent of evolution under punctuated equilibrium. A simulation under constant population conditions shows the dynamics of takeover and extinction of different composomes, leading to the generation of different population distributions. The P-GARD model offers a scenario whereby biopolymer formation may be a result of rather than a prerequisite for early life-like processes.

1   Introduction

Contemporary life is strongly based on biopolymers, serving as structural elements, for catalysis and signaling pathways, as well as for genetic information storage (Stryer 1995; Lodish *et al.* 1999; Alberts *et al.* 2002; Gomperts *et al.* 2002). Though simpler molecules could perform some of these functions, e.g. catalysis (Kochavi *et al.* 1997; Suwannachot and Rode 1998; Mitsuzawa and Watanabe 2001), biopolymers prevail in present-day life because of their better effectiveness and specificity.

A point of view, which is hardly disputed, is that organic synthesis is a prerequisite for life. Numerous energy sources and reaction pathways have been explored that lead to the terrestrial formation of relatively small organic compounds (Miller 1986), e.g. amino acids (Miller 1953), carbohydrates (Cooper *et al.* 2001), nucleotide bases (Basile *et al.* 1984; Eschenmoser 1999) and lipids (Ourisson and Nakatani 1994). Many of these compounds may have in parallel supplied from extraterrestrial sources such as dust particles, meteorites, and comets (Briggs and Mamikunian 1964; Lawless 1980; Anders 1989; Deamer and Pashley 1989; Oro *et al.* 1992; Maurette 1998; Sephton 2002). Such simple organic compounds, often termed 'monomers', are believed to have constituted the building blocks for the more complex molecules necessary for life. However, a widely accepted school of thought asserts that the more complex compounds, typically linear polymers of amino acids or of nucleotides, were prerequisites for the emergence of life. Accordingly, it is necessary to invoke abiotic mechanisms, e.g. mineral catalysis, that could allow polymer formation (Ferris *et al.* 1979; Acevedo and Orgel 1986; Ferris and Ertem 1993; Ertem and Ferris 1997; Bujdak and Rode 1999).

Several authors, including our own group, have proposed an alternative point of view. This contends that certain life-like processes could have emerged *prior* to the synthesis of biopolymers (Morowitz H.J. *et al.* 1988; Wachtershauser 1988; Bagley and Farmer 1991; Segre *et al.* 2000; Segre and Lancet 2000; Szathmary 2000; Shenhav *et al.* 2003). Such life-like entities may have been driven by mechanisms of mutual catalysis among constituent organic molecules (Kauffman 1993; Stadler *et al.* 1993; Dyson 1999; Jain and Krishna 2001; Kaneko 2002). Furthermore, according to this school of thought, biopolymers could be the result of primitive selection and rudimentary evolution rather than prerequisites for them. This view also suggests that the prebiotic arena consisted of a very diverse monomer repertoire, a random chemistry scenario (Segre and Lancet 1999; Wills and Bada 2001), and that polymerization was one of the mechanisms that afforded the selection of the monomer sub-repertoire that became involved in more elaborate life forms.

We propose here a model, which delineates a quantitative pathway for oligomer formation, as well as for monomer selection. In this model, mesobiotic entities (Shenhav *et al.* 2003) with the capacity to replicate, are envisioned as bridging the gap between a prebiotic 'monomer world' (Shapiro 1986) and more elaborate biotic forms that, like contemporary cellular life, are based on DNA, RNA and proteins. We suggest that such mesobiotic entities could serve as cradles for endogenous biopolymer synthesis.

2   Results

2.1   DESCRIPTION OF THE P-GARD MODEL

As in the previously described Graded Autocatalysis Replication Domain (GARD) model (Segre *et al.* 1998a; Segre *et al.* 2000; Segre and Lancet 2000; Segre *et al.* 2001a) the presently elaborated Polymer GARD (P-GARD) depicts the chemical

dynamics of primordial molecular assemblies. Based on a random chemistry scenario, P-GARD considers a repertoire of N_G different molecular monomer types. For the sake of chemical realism, but without the loss of generality, these monomers are assumed to be amphiphilic, and thus to spontaneously form non-covalent assemblies. As in the GARD model, a simplified Oparin/Morowitz-style (Oparin 1953; Oparin 1957; Morowitz H.J. 1967) progeny-generation process is assumed, in the course of which simple physical fission (or split) results in the transfer of compositional information (Segre *et al.* 2000; Segre *et al.* 2001b). Homeostatic growth, sustaining relatively stable compositions, affords numerous cycles of progeny generation with continuity of a 'compositional genome' (Segre *et al.* 2000).

In the monomer-only GARD model, mutual catalysis was invoked as modulating the basal forward (entry) and backward (exit) rate constants (k_f, k_b) by some of the monomeric molecules within the assembly. The amount of mutual catalysis (β_{ij}) was derived from a statistical chemistry model and from accompanying experimental data (Lancet *et al.* 1993; Segre *et al.* 2001a; Segre *et al.* 2001b; Rosenwald *et al.* 2002), suggesting that the rate enhancement values could be approximated by a lognormal distribution.

Here we extend the basic GARD model by introducing an additional type of chemistry, namely the endogenously catalyzed formation of covalent bonds. In future elaborations of P-GARD, oligomers of different length will be considered. Here, in order to examine the most basic consequences of oligomer formation, the monomers are assumed to be monovalent, hence only dimers can form (Fig. 1). Thus, P-GARD includes reactions of the form:

$$A_i + A_j \;\rightleftharpoons\; A_i\text{-}A_j \quad (1)$$

where A_i and A_j are arbitrary molecular species present in the assembly and A_i-A_j is their covalent dimer. This could be viewed, in simplified chemically realistic terms, as

two amphiphiles, having a polar head and a lipophilic tail, joining in a head-to-head manner, but the model should be taken to represent other possibilities, e.g. the involvement of the tail groups in the formation of a covalent bond, as well as the linking of an isolated polar head group to an amphiphile head group, generating a double-headed amphiphile. Potential chemical reactions underlying such covalent reactions could be ester, thioester, aldol, anhydride amide or phosphoester condensations. Also implicitly assumed is some form of activation that will allow the down-hill formation of covalent bonds. However, in order to maintain the generality of the analysis we refrain here from specifying the precise chemistry.

We denote the spontaneous forward and backward rate constants of the covalent reactions in Eq. 1 by k_{poly} and k_{break} respectively. For simplicity, and as previously described (Segre *et al.* 2000; Segre *et al.* 2001a) these rate constants are taken to be equal for all molecular species. A non-specific entropy-related enhancement factor resulting from increased local concentrations of compounds within the assembly is accounted for (Fendler and Fendler 1975). As in the basic GARD model, we invoke catalysis exerted on covalent bond formation and dissociation, whereby the catalyzing molecules may be either monomer or dimers. The catalytic coefficients specifying the degree to which a compound X catalyzes the covalent association of monomers A_i and A_j is denoted by $\gamma_{A_i,A_j,X}$.

Like in the simpler monomer GARD, values of the catalytic enhancements exerted by monomers on the join/leave reactions of other monomers are obtained using a statistical chemistry approach. To avoid an oversized parameter lookup table, and in order to capture likely correlations between the catalysis events exerted by more complex molecules and those effected by their constituent monomers, the catalytic rates enhancements that involve dimers are calculated by a formula rather than being drawn from a distribution. We have used the sum of the catalysis exerted

6

by the monomers composing the dimers (Fig. 1), though other functions could be considered. For example, the catalysis exerted on the join/leave reactions of the molecular species A_i by a dimer A_j-A_k inside the assembly would be:

$$\beta_{Ai,Aj-Ak} = \beta_{Ai,Aj} + \beta_{Ai,Ak} \quad (2)$$

where the second index denotes the catalyst. Similarly, the rate enhancement effect on bond formation/break reaction, exerted by a monomer A_k on the covalent reaction between monomeric species A_i and A_j may be taken to be

$$\gamma_{Ai-Aj,Ak} = \beta_{Ai,Ak} + \beta_{Aj,Ak} \quad (3)$$

For the case in which a dimer A_k-A_l catalyses the formation of another dimer A_i-A_j, we have assumed independent tandem reactions, in a templating-like fashion (Fig. 1)

$$\gamma_{Ai-Aj,Ak-Al} = \beta_{Ai,Ak-Al} + \beta_{Aj,Ak-Al} = \beta_{Ai,Ak} + \beta_{Ai,Al} + \beta_{Aj,Ak} + \beta_{Aj,Al} \quad (4)$$

Thus, a total of N_G^2 covalent bond formation reactions are added to the basic N_G join/leave reactions of the basic GARD model. The number of rate enhancement factors is thus $(N_G^2+N_G)^2$, as compared to N_G^2 for monomer joining GARD.

2.2   P-GARD SIMULATIONS

Fig. 2 shows a time autocorrelation diagram of a dynamic behavior resulting from the P-GARD model. This diagram shows compositional similarity (hue) as well as degree of dimerization (saturation). Red colored squares that indicate preservation of chemical composition across several assembly fission events are indicative of the occurrence of quasi-stationary states - composomes (Segre *et al.* 2000), and these are seen to contain different molar fraction of dimers. Of particular interest are long lasting composomes such as the one seen between time points 500 and 750, with relatively high concentration of dimers – a situation that may be considered a 'dimer takeover'. Whereas only monomers are available externally, dimers are synthesized *in situ,* within the assembly. As for monomer-based GARD (Segre *et al.* 2000), the dynamics is strictly dependent on the fact that the interior of the assembly is kept

7

away from equilibrium by repeated splits and infinite external supply of monomers (constant concentration of monomers in the environment). It is generally observed that the fraction of dimers within the assembly increases with the covalent bond-related basal rate constants, k_{poly} and k_{break} (data not shown).

When different sets of catalytic enhancement values (β_{ij}) are used, even when based on the same underlying statistical distribution, different dynamic behaviors obtain (Fig. 3). In some simulation runs well-defined composomes are observed (R_1), in other cases the assembly appears more inclined towards random drift, and stable compositions are not reached (R_2) and in yet other cases (R_3) a single dynamically stable composition prevails.

In order to gain more quantitative measures on the global dynamic behavior of P-GARD and on the modes of its underlying emergence of composomes, we have applied an 'on-the-fly' clustering algorithm (Fig. 4). The algorithm tags compositions just prior to a fission event as belonging to a group with sufficient internal similarity, so as to be classified as belonging to the same composomal cluster. Fig 5A show the result of the clustering procedure as applied to the time evolution of a single run (R_1). Typically, composomes that appear early in the simulation tend also to belong to the more highly populated composomal clusters, but this correlation is not strict. Thus, some clusters which first appear late in the trace still reappear rather frequently (e.g. composome c20).

The different composomes are distinct, but may have overlapping compositions (Fig. 5B). In some cases the degree of compositional similarity is very low, i.e. the composomes are orthogonal. While the number of such orthogonal composomes is limited, the rate of appearance of new composomes appears to obey a time-dependent power law (Fig. 5C), similar to the behavior of a system evolving under punctuated equilibrium (Bak and Sneppen 1993; Root-Bernstein and Dillon 1997; Segre *et al.*

1998b; Jain and Krishna 2002). This implies an 'open ended' behavior, whereby, at least over a broad range of time periods, additional composomes would continue to appear as the simulation progresses.

Sample detailed pathways leading to the synthesis of dimers within a quasi-stationary-state assembly are portrayed in Fig. 6. Although every reaction is catalyzed by all molecular species, some of catalysis values are significant and the others are negligible. Distinct dimer compositions result from different underlying metabolism-like networks, where different monomer and dimer constitutions significantly catalyze the joining of selected monomers and the synthesis of selected dimers. Some of the catalyzed species in turn serve as catalyst for additional reactions in a mutually catalytic fashion. Other molecular species are formed in a pathway aided by catalysis, but do not exert significant rate enhancement on other compounds, and thus constitute 'parasites' (Fig. 6A). Networks corresponding to very stable compositions (e.g. R_3 in Fig. 3) tend to consist of a relative small molecular repertoire (not shown).

We asked whether a correlation exists among global parameters of the dynamics of P-GARD in different simulation runs. One hundred simulations (all with the same basal rate constants but different β were carried out with different β_{ij} values, each lasting 200,000 time units. Three global parameters were analyzed (Fig. 7A, Fig. 8A, B): a) the fraction of time spent in composomes, as compared to drift mode; b) the number N_c of distinct "newly discovered" composomes that appear up to t=200,000; c) the dimer content of each composome. A clear trend was discerned, whereby, up to a point, N_c was larger in runs in which the assembly spent a large fraction of the time in composomal states. However, the highest values of the fraction of time spent in composomes were associated with runs in which N_c was small, cases in which one or very few dynamically stable composomes appeared. Appreciable dimer content began

9

to appear when the fraction of time spent in composomes was higher than 15%, and reached high values above the value of 30%.

2.3 THE EFFECT OF INHIBITION

The Basic GARD and P-GARD models are assumed to consist of molecules that can only exert positive rate changes, while the possibility of rate diminution was not allowed, for the sake of simplicity. Inhibition would amount to populating the β matrix with negative values, intermixed with positive ones. Fig. 7B shows the results of 100 simulations identical to those described in Fig. 7A, but in which a randomly selected half of the rate modifying values were multiplied by -1. The three global parameters discussed in the previous section were compared between the two sets of simulations. They exhibited the same distributions with minor differences. We found a somewhat higher fraction of time spent in composomes and broader distribution for the fraction of dimers (Fig. 8A, B), as well as a lower average count of different composomes (not shown). The general similarities in the global parameters between the two simulation sets confirm the notion that a simplified model with no inhibition is acceptable. This similarity may be rationalized as follows: because only a few dominant rate enhancement values occur for each species, and because these values are log-normally distributed, the probability of negating a dominant rate enhancement is not very large. Furthermore, even if such events occurred, this would occasionally eliminate a component of a relatively complex composomal network, and this should have a small effect on the global behavior of the GARD simulation.

2.4 P-GARD POPULATION DYNAMICS

To demonstrate that composomes are promising higher-level entities they should display mutual competition. In the foregoing simulations, a GARD "track" was followed, in which after a split, one of the two progeny was selected at random (Segre *et al.* 2000; Segre *et al.* 2001b). This prevented us from analyzing the complex

10

interdependence between assemblies potentially existing in a population of GARD assemblies. We therefore conducted a preliminary simulation in which assemblies in different composomal states co-exist under constant population conditions. In these simulations, composome fitness appears as an emergent property.

A population-related simulation is based on parameters derived from a given GARD track with N_c different composomal states (composomes). For every composomal state C_i, three emergent fitness parameters are calculated by from a GARD simulation of a single assembly by tracking all the composomes encountered along the simulation. The drift state C_0 is considered as a virtual composome in the sense that all parameters can be calculated for it as well. The parameters are:

T_i - the typical growth time of the composome, i.e. the time elapsing between assembly splits.

S_i - the probability of the progeny to survive a split, i.e. preserving its parental composomal state.

E_i - the probability of entering the composomal state i after a split from a different composome.

An initial population of assemblies is seeded with q copies of each of the N_c composomes. For an assembly in the composomal state C_i, the time of upcoming split is given by T_i and the two progenies are assigned composomal states by one or two consecutive "drawings", first, using S_i a decision is made about remaining in the state C_i, and if it fails, then all E_j (j≠i) are used for assigning a transition to a state C_j. Every split is coupled with a random death event, fulfilling the constant population constrain. The population simulation is conducted under the following assumptions: a) no memory – the probability of entering another composomal state is independent of the parental composomal state; b) progenies independence – the composomal state assignments of two progeny assemblies are independent of each other;

11

Fig. 9A,B illustrates a population dynamics simulation based on a track with 33,000 splits using the rate enhancement values of simulation run R_1 of Fig. 7A, with 77 composomal states. It is seen that only five of these survive to an appreciable extent, while all others diminish to very low representation or diminish altogether. The remaining composomes undergo a complex dynamics, including oscillations at various frequencies, with different ones becoming dominant at different time periods.

4 Discussion

Different basic rate enhancements values (β_{ij}) may be taken to represent different prebiotic mixtures containing different global chemistries. The foregoing quantitative analysis of 100 different random selections of catalytic values indicates the existence of intermediate cases between those displaying mainly random drift and those with very few stable composomal states. Some composomes in these intermediates also contain an appreciable number of dimers. These are deemed more significant in the context of understanding prebiotic evolution.

The basic GARD model considers only join/leave reactions whereby the synthesis of all molecules concerned takes place external to the assembly. Here we have shown that if we consider also reactions of formation of novel molecules (oligomers), dynamic assemblies emerge, which contain significant amounts of complex molecules, not available externally. In other words, the molecular assembly possesses the capacity to select proper monomers from the environment and then synthesize selected composite molecules from such precursors.

Autotrophy, based solely on the input of inorganic molecules, was considered to be the essence of life emergence (Wachtershauser 1988; Maden 1995; Lazcano and Miller 1999). Yet, in light of the successes in prebiotic synthesis of organic compounds and the availability of such molecules from extraterrestrial sources (Miller

1986; Oro *et al.* 1992) this requirement may be considered less strictly necessary. Thus, the capacity for selective simulated synthesis of more complex organic compounds form simpler ones by the P-GARD model is of significance. These simulations help establish the potential existence of mesobiotic entities (Shenhav *et al.* 2003) that may have formed a bridge between the prebiotic synthesis of monomeric organic compounds and the emergence of more elaborate life-like entities that contained biopolymers.

This study, similar to previous ones (Bagley and Farmer 1991; Root-Bernstein and Dillon 1997; Segre *et al.* 2000; Jain and Krishna 2001) puts emphasis on the elucidation of potential origin of life processes related to spontaneous molecular organization, stemming from the random chemistry scenario likely to have prevailed on primordial earth. We consider the formation of non-covalent molecular ensembles to be a key process. Previously, we have shown that mutual catalysis within such assemblies provides the capacity to transfer compositional information to progeny (Segre *et al.* 2000; Segre and Lancet 2000; Segre *et al.* 2001b). Here we have shown that the capacity to transfer such information is retained even when more complex molecules and a much larger repertoire are considered.

Previous models have invoked mutual catalytic sets of oligomeric molecules (Bagley and Farmer 1991; Kauffman 1993; Dyson 1999). In some of these examples, the model consisted of two or relatively few monomer types, and considered a very large number of derived oligomers for the establishment of catalytic closure. Criticism has been raised indicating that diminishing likelihood of realistic molecular sets that contained a number of molecules of the order of the magnitude of the entire possible oligomer repertoire. Such a problem is significantly alleviated in the P-GARD model, in which small mutual catalytic subsets are generated. This is despite the fact that a much larger number of monomer types is included in the reaction scheme.

13

The capacity of P-GARD assemblies to manifest small closed mutually catalytic sets is due to the metabolism-like structure capable of selection of a small sub-repertoire of monomers out of all available, combined with an even stricter selection of an oligomer sub-repertoire out of all those formally possible. In more complex future embodiments of P-GARD, in which higher order oligomers would be allowed, it is anticipated that such 'pruning' effect would become even more crucial. This will mimic the state of affairs in present-day cellular life, where only an infinitesimally small fraction of all possible oligomers and polymers actually get synthesized. While observing in the test tube the elaborate chain of events that led to such strict selection may not be straightforward, computer simulations of the type reported here could allow at least limited insight into the nature of such crucial processes.

An interesting feature of GARD, further enhanced in P-GARD, is its potential capacity to withstand putative side reactions and parasitic molecular species, as well as its including "keystone"-like species, as previously shown in analogous simulated networks (Jain and Krishna 2002). Similarly, our system exhibits robustness even in the presence of large fractions of parasite-like species. In addition, as previously described by the same group (Jain and Krishna 2001) our present simulations suggest that mutually inhibitory species within a catalytic network are not strongly detrimental to the network integrity and dynamics.

Evolutionary processes of replicating metabolic networks have been simulated by the imposed introduction of new molecular species on top of a pre-existing repertoire to effect mutation-like changes (Bagley *et al.* 1991; Segre *et al.* 1998b; Jain and Krishna 2002). P-GARD, particularly when extended in the future to higher order oligomers, harbors a built-in mechanism for producing a very large number of novel molecular species as part of its endogenous chemistry. Furthermore, under the regimen of endogenous emergence of novel dimer-containing compositions, a time-

dependent power law is obeyed, similar to that previously seen in other simulated evolving dynamic systems (Bak and Sneppen 1993; Segre *et al.* 1998b) governed by externally-imposed mutations. Thus P-GARD may make it possible to observe a more natural open-ended evolution like dynamics.

The present simulation of GARD assembly populations is different from that previously published by our group (Segre *et al.* 2000). In the former analysis only a very limited number of assemblies could be followed because of computing power constraints, as the dynamics of each assembly was simulated in full molecular detail. Here, a phenomenological formalism was utilized, whereby each assembly was represented by a set of three emergent fitness parameters, allowing us to handle hundreds of assemblies for thousands of generations. This high-level treatment is a promising avenue for future extensions to longer polymers.

The introduced notion of fitness parameters also portrays the composome population as analogous to a quasi-species (Eigen 1971; Eigen *et al.* 1988), where each composome is analogous to a replicating polynucleotide with its own reproductivity (T_i), and error rate related to (1- S_i). The dynamics of our system also shows a progression similar to natural selection, where only a few composomes exhibit non negligible quantity in the population, while the other may be considered as extinct.

A point of criticism against a 'metabolism first' scenario (Morowitz H.Z. 2002) is its lack of heredity and evolutionary potential (Szathmary 2000). Metabolism-like networks such as depicted by the P-GARD model, and similar to other autocatalytic sets may help alleviate such criticism. The fact that P-GARD is capable of sustaining a specific composition of dimers through a series of growth and split events, and also that in population simulations it shows extinctions and survivals, signifies

characteristics resembling evolution, thus helping to bridge the dual requirement for
metabolism and for the propagation of information.

References

Acevedo, O.L. and Orgel, L.E.: 1986, Template-directed oligonucleotide ligation on
hydroxylapatite *Nature*, **321**, 790-792

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P.: 2002,
Molecular Biology of the Cell, Garland Pub,

Anders, E.: 1989, Pre-biotic organic matter from comets and asteroids *Nature*, **342**,
255-257

Bagley, R.J. and Farmer, D.J.: 1991, Spontaneous Emergence of a Metabolism in
Langton, C. G., Taylor, C., Farmer, J. D. and Rasmussen, S.(eds), *Artificial
Life II*, Addison-Wesley, Redwood City, pp. 93-140

Bagley, R.J., Farmer, D.J. and Fontana, W.: 1991, Evolution of a Metabolism in
Langton, C. G., Taylor, C., Farmer, J. D. and Rasmussen, S.(eds), *Artificial
Life II*, Addison-Wesley, Redwood City, pp. 141-158

Bak, P. and Sneppen, K.: 1993, Punctuated Equilibrium and Criticality in a Simple
Model of Evolution **71**, 4083-4086

Basile, B., Lazcano, A. and Oro, J.: 1984, Prebiotic syntheses of purines and
pyrimidines *Adv Space Res*, **4**, 125-131

Briggs, M.H. and Mamikunian, G.: 1964, Organic constituents of carbonaceous
chondrites *Life Sci Space Res*, **2**, 57-85

Bujdak, J. and Rode, B.M.: 1999, The effect of clay structure on peptide bond
formation catalysis *J. Mol. Catal. A-Chem.*, **144**, 129-136

Cooper, G., Kimmich, N., Belisle, W., Sarinana, J., Brabham, K. and Garrel, L.: 2001,
Carbonaceous meteorites as a source of sugar-related organic compounds for
the early Earth *Nature*, **414**, 879-883.

Deamer, D.W. and Pashley, R.M.: 1989, Amphiphilic components of the Murchison carbonaceous chondrite: surface properties and membrane formation *Orig Life Evol Biosph*, **19**, 21-38

Dyson, F.: 1999, *Origins of life*, Cambridge University Press, Cambridge.

Eigen, M.: 1971, Selforganization of Matter and Evolution of Biological Macromolecules *Naturwissenschaften*, **58**, 465-&

Eigen, M., McCaskill, J. and Schuster, P.: 1988, Molecular Quasi-Species *J. Phys. Chem.*, **92**, 6881-6891

Ertem, G. and Ferris, J.P.: 1997, Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds *J Am Chem Soc*, **119**, 7197-7201

Eschenmoser, A.: 1999, Chemical etiology of nucleic acid structure *Science*, **284**, 2118-2124

Fendler, H.J. and Fendler, E.J.: 1975, *Catalysis in Micellar and Macromolecular Systems*, Academic Press, New York.

Ferris, J.P., Edelson, E.H., Mount, N.M. and Sullivan, A.E.: 1979, The effect of clays on the oligomerization of HCN *J Mol Evol*, **13**, 317-330

Ferris, J.P. and Ertem, G.: 1993, Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA *J Am Chem Soc*, **115**, 12270-12275

Gomperts, B., Kramer, I.M., Tatham, P.E.R. and Gomperts, B.D.: 2002, *Signal Transduction*, Academic Press,

Jain, S. and Krishna, S.: 2001, A model for the emergence of cooperation, interdependence, and Structure in evolving networks *P Natl Acad Sci USA*, **98**, 543-547

Jain, S. and Krishna, S.: 2002, Large extinctions in an evolutionary model: The role of innovation and keystone species *P Natl Acad Sci USA*, **99**, 2055-2060

Kaneko, K.: 2002, Kinetic origin of heredity in a replicating system with a catalytic network *J. Biol. Phys.*, **28**, 781-792

Kauffman, S.: 1993, *The origin of order*, Oxford University Press,

Kochavi, E., Bar-Nun, A. and Fleminger, G.: 1997, Substrate-directed formation of small biocatalysts under prebiotic conditions *Journal of Molecular Evolution*, **45**, 342-351

Lancet, D., Sadovsky, E. and Seidemann, E.: 1993, Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system *Proc Natl Acad Sci U S A*, **90**, 3715-3719

Lawless, J.G.: 1980, Organic compounds in meteorites *Life Sci Space Res*, **18**, 19-27

Lazcano, A. and Miller, S.L.: 1999, On the origin of metabolic pathways *J Mol Evol*, **49**, 424-431

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J. and Zipursky, L.: 1999, *Molecular Cell Biology*, W H Freeman & Co.,

Maden, B.E.: 1995, No soup for starters? Autotrophy and the origins of metabolism *TIBS*, **20**, 337-341

Maurette, M.: 1998, Carbonaceous micrometeorites and the origin of life *Orig Life Evol Biosph*, **28**, 385-412

Miller, S.L.: 1953, A production of amino acids under possible earth conditions *Science*, **117**, 528-529

Miller, S.L.: 1986, Current status of the prebiotic synthesis of small molecules *Chem Scr*, **26B**, 5-11

Mitsuzawa, S. and Watanabe, S.-i.: 2001, Continuous growth of autocatalytic sets *Biosystems*, **59**, 61-69

Morowitz, H.J.: 1967, Biological self-replicating systems in Snell, F. M.(ed) *Progress in Theoretical Biology*, Academic Press, pp. 35-58

Morowitz, H.J., Heinz, B. and Deamer, D.W.: 1988, The chemical logic of a minimum protocell *Origins Life Evol B*, **18**, 281-287

Morowitz, H.Z.: 2002, *The Emergence of Everything*, Oxford University Press, New York.

Oparin, A.I.: 1953, *The origin of life*, Dover Pub., N.Y.

Oparin, A.I.: 1957, *The origin of life on the earth*, Oliver and Boyd, London.

Oro, J., Mills, T. and Lazcano, A.: 1992, The cometary contribution to prebiotic chemistry *Adv Space Res*, **12**, 33-41

Ourisson, G. and Nakatani, Y.: 1994, The terpenoid theory of the origin of cellular life: the evolution of terpenoids to choloesterol *Chemistry & Biology*, **1**, 11-23

Root-Bernstein, R.S. and Dillon, P.F.: 1997, Molecular complementarity .1. the complementarity theory of the origin and evolution of life *Journal of Theoretical Biology*, **188**, 447-479

Rosenwald, S., Kafri, R. and Lancet, D.: 2002, Test of a statistical model for molecular recognition in biological repertoires *J Theor Biol*, **216**, 327-336

Segre, D., Ben-Eli, D., Deamer, D.W. and Lancet, D.: 2001a, The lipid world *Origins Life Evol B*, **31**, 119-145

Segre, D., Ben-Eli, D. and Lancet, D.: 2000, Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies *P Natl Acad Sci USA*, **97**, 4112-4117

Segre, D. and Lancet, D.: 1999, A statistical chemistry approach to the origin of life *Chemtracts - Biochemistry and Molecular Biology*, **12**, 382-397

Segre, D. and Lancet, D.: 2000, Composing life *Embo Rep*, **1**, 217-222

Segre, D., Lancet, D., Kedem, O. and Pilpel, Y.: 1998a, Graded autocatalysis replication domain (GARD): Kinetic analysis of self-replication in mutually catalytic sets *Origins Life Evol B*, **28**, 501-514

Segre, D., Pilpel, Y. and Lancet, D.: 1998b, Mutual catalysis in sets of prebiotic organic molecules: Evolution through computer simulated chemical kinetics *Physica A: Statistical and Theoretical Physics*, **249**, 558-564

Segre, D., Shenhav, B., Kafri, R. and Lancet, D.: 2001b, The molecular roots of compositional inheritance *J Theor Biol*, **213**, 481-491

Sephton, M.A.: 2002, Organic compounds in carbonaceous meteorites *Nat Prod Rep*, **19**, 292-311

Shapiro, R.: 1986, *Origins: A skeptic's guide to the creation of life on Earth*, Simon & Schuster, Inc., N.Y.

Shenhav, B., Segre, D. and Lancet, D.: 2003, Mesobiotic emergence: molecular assemblies that self-replicate without biopolymers *Adv Complex Sys*,

Stadler, P.F., Fontana, W. and Miller, J.H.: 1993, Random catalytic reaction networks **63**, 378-392

Stryer, L.: 1995, *Biochemistry*, W H Freeman and Co,

Suwannachot, Y. and Rode, B.M.: 1998, Catalysis of dialanine formation by glycine in the salt-induced peptide formation reaction *Orig Life Evol Biosph*, **28**, 79-90

Szathmary, E.: 2000, The evolution of replicators *Philos Trans R Soc Lond B Biol Sci*, **355**, 1669-1676

Wachtershauser, G.: 1988, Before Enzymes and Templates: Theory of Surface Metabolism *Microbiol. Rev.*, 452-484

Wills, C. and Bada, J.: 2001, *The Spark of Life: Darwin and the Primeval Soup*, Perseus Publishing,

Legends

A cartoon depicting reactions and catalytic processes considered by P-GARD. Different colored balls represent different monomeric compounds A_i. The dimerization reaction proceeds by way of template like catalysis, with the transition state complex closed in brackets. The interactions within such complex are assumed to resemble those that underlie rate enhancement of monomer entry/exit to an assembly.

A 'compositional carpet' of the P-GARD simulation run R_1 (figure 7A). The drawing depicts a time correlation matrix, where both the ordinate and the abscissa represent the same time scale in arbitrary simulation units. Each point on the 2-dimensional diagram represents how similar are the compositions at the two time points compared. Red colors signify high similarity (H nearly 1) and blue colors suggest low similarity (H nearly 0). The similarity measure that was used is the dot product of the normalized samples (Segre *et al.* 2000):

$$H(\mathbf{S}_{t1}, \mathbf{S}_{t2}) = \frac{\mathbf{S}_{t1} \cdot \mathbf{S}_{t2}}{|\mathbf{S}_{t1}||\mathbf{S}_{t2}|}$$

where $\mathbf{S}_i$ is a compositional vector and t1 and t2 are time points. The brightness of the color indicates the total fraction of dimers within the compared composition. Vivid colors relate to high levels of dimers, pale relate to high levels of monomers (bottom panel).

The saw-tooth graph below the autocorrelation diagram depicts the size of the assembly in terms of molecular count (N) as a function of time. We have assumed a

stochastic fission process when N was greater than 80. After a fission event, the simulation follows one randomly selected progeny assembly out of the two generated. Typically, a split occurs on average every 40-60 simulation time points.

The repertoire size taken in this simulation is $N_G = 100$. The parameters governing the chemical distribution of β are $\mu = -4$, $\sigma = 4$. The spontaneous reaction rates (k_f, k_b, k_{poly} and k_{break}) where all taken to be equal to one. The concentrations of monomers outside of the assembly are kept constant at a value of 1000 per unit volume, while the internal scaled concentrations are the molecular counts.

Figure 3

The compositional 'carpets' of three different simulation runs (R_1, R_2, R_3, as marked in figure 7A) shown at four different time scales. These are conducted with the same governing statistical parameters as in figure 2, but having different specific values populating the β matrix. Three dynamic behaviors are depicted: R_1, the emergence of semi-stationary composition consisting of high level of dimers; R_2, mainly random drift, with moderate level of dimers; R_3, a system governed by a strong dimer autocatalyst.
While R_1 shows interesting behavior with a diversity of composomal states, R_2 is unable to sustain compositional information while R_3 does not posses adequate diversity needed for a capacity to evolve.

Figure 4

A flow chart of the 'on-the-fly' clustering algorithm. The input of the algorithm is a time trace of sampled compositions just before fission, and the outputs are tagged sub-traces. Assembly compositional states not sufficiently similar to the previous (parental) tagged sample are considered to be non-homeostatic, i.e. in a drift period

and are tagged as 'cluster' C_0. Those sufficiently similar to the parental assembly are considered to be in a homeostatic quasi-stationary state and are tagged with the most similar existing cluster, if the similarity to this cluster is above a threshold else are taken to seed a new cluster. The threshold value used was H=0.8.

Figure 5

Results of the 'on-the-fly' clustering algorithm when applied to a single simulation run (R_1 in figure 7A).

A. The abscissa indicates consecutive splits. c0 to c29 refer to the different composomal clusters (see figure 4). Each cluster is summarized by a circle on the right ordinate, whose size reflects frequency of the composome's appearance and color indicates the dimer content as shown in the scale on the right.

B. The chemical composition of clusters C1 to C29. Columns represent different clusters, rows different molecular species. Molecular species which were negligible (below 1%) on all clusters are not shown. Species marked with 2 digit values representing monomers, and four digit values corresponding to dimers made of the indicated to monomers.

C. The linear power law dependence of the number of different composomes emerging (N_c) as a function of the number of splits that an assembly undergoes (Sp). The red line derives from the simulation of figure 5A, the black line is a differently randomized and 10 fold longer simulation with the same β_{ij} parameters. The red line obeys $N_c = 1.419\ Sp^{0.387}$ while the black line has the equation $N_c = 0.323\ Sp^{0.528}$.

Figure 6

The networks corresponding to the composomal clusters C3 and C7 in figure 5B.

A. A schematic representation of the 'metabolic' mutually catalytic networks that underlie the two composomes. The scheme describe enumerated monomers (on the left) indicated by different arbitrary colors with rounded rectangles representing external disposition while rectangles indicate being within an assembly. Dimers are indicated by the colors of the constituent monomers. The Catalysts on the join/leave reactions and bond formation/break reactions are indicated by the shapes above the relevant reaction link. The percentage indicates the fraction of the compound in the assembly's composition. Some of the molecular species (e.g. 9 and 89 in C3 or 88 in C7) are catalyzed but do not contribute any significant catalysis and thus could be considered to be 'parasites'.

B. Another view of the above networks as described in previous publications (Segre *et al.* 2000). Molecules are marked by green circle while catalysis with blue arrows. As in figure 5, species marked with values up to 100 represent monomers, four digit values represent dimers of the two-digit enumerated monomers. Catalysis exerted on monomers refers to join/leave reaction and catalysis exerted on dimers refers to bond formation/break reaction.

Figure 7

A. A summary of the clusters encountered in 100 different runs taken with the same parameters as described in figure 2, except those listed below. The runs are represented by columns, ordered by the total amount of time spent in a quasi-stationary composition, indicated on the abscissa and on top of each column. The

runs differ solely by the catalytic rate enhancement matrices (β), though their values distribution complies with the same statistical chemistry constraints (μ = -4, σ = 4). The number of different composomes generated up to time point 200,000 is N_c and the color code and circle sizes are as in figure 5A. In 85% of the runs, the assembly was mainly (> 65%) in a drift mode, while in 7% the assembly exhibited strong tendency (>90% of the time) to hold a dynamically stable composition.

B. The effect of inhibition. A summary of the clusters encountered in 100 simulations conducted with the same catalytic rate enhancement matrices as in A, in which a randomly selected half of the rate value were multiplied by -1, turning them from rate enhancement into an inhibitory effect. The minimal rate for any single reaction (accounting for all positive and negative modulations) is naturally limited to zero, as negative chemical rates do not have a physical meaning. All other simulation parameters and the presentation of the results are as in A. The behavior has a considerable similarity to that observed in A, as quantified in Fig. 8.

Figure 8

Comparison of global parameters calculated in the two sets of simulations described in figure 7. The black bars indicate the set of simulations conducted without inhibition and the gray bars are for the set of simulation done with inhibition.

A. Binning of the simulations according to the fraction of time spent in composomal states. The two distributions share the same characteristics, though in simulations with inhibition assemblies tend to spend more time in composomal states as

compared to without inhibition (averages 26.6% and 18.1% respectively, P= 0.0311. Wilcoxon rank sum test).

B. Binning of the composomes encountered in all simulations according to dimer fraction. When simulations with and without inhibition are compared, the mean dimer fractions are almost identical (46% vs. 45%) while the standard deviation slightly differ (14.6% vs. 12.7%) (P= 0.0406, Wilcoxon rank sum test).

Figure 9

Population dynamics simulations for GARD assemblies, with computed fitness parameters of the composomes encountered in a GARD simulation with the same β matrix as in simulation R1 in figure 7A. Each color (blue, green, red, cyan and magenta) represents an assembly in one of the five major (most frequent) composomal states.

A. An illustrative sample simulation, whereby the initial population consisted of a single copy (q=1) of each of the five major (most frequent) composomes. Yellow circles indicate the rare appearances of minor composomal states, and colorless circles are in drift mode. Each column represents a time point in which one of the assemblies underwent fission. In each time point an event of assembly death also occurred, marked by X.

B. A simulation of population dynamics of all N_c=77 composomes with parameters as described in A, and with initial population of q=3 for each. The changes in the counts of the five major composomes are shown. Other composomal states, including the drift rapidly diminished or disappeared (not shown). Three different periods, delimited by broken lines, may be discerned: a first transient period in which a sharp rise in the quantities of the major

composomes occurred, concomitant with the disappearance of other composomes; a second and a third period, that portray fluctuations of the major composomes around quasi-stable quantities, with different ratios for each of the periods. The dimer content of the five major composomes was: Blue, 0.82; Green, 0.43; Red, 0.8; Cyan, 0.56; Magenta, 0.79.
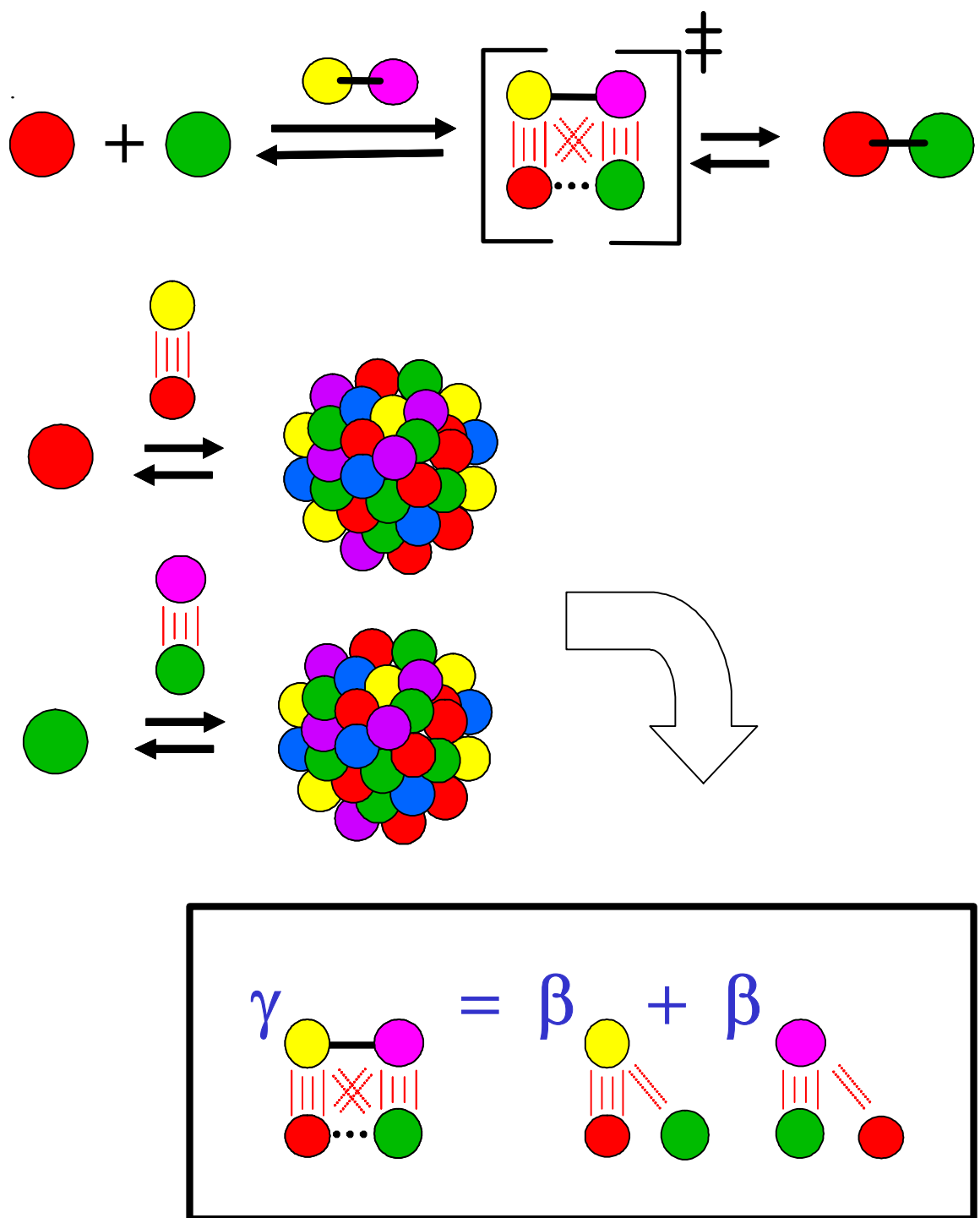
Figure 1

Figure 2

Figure 3

Trace

No ← Split ? → Yes

Yes ← Similar to Parents ? → No

No ← Known cluster ? → Yes

Add a cluster

Add to cluster

Done
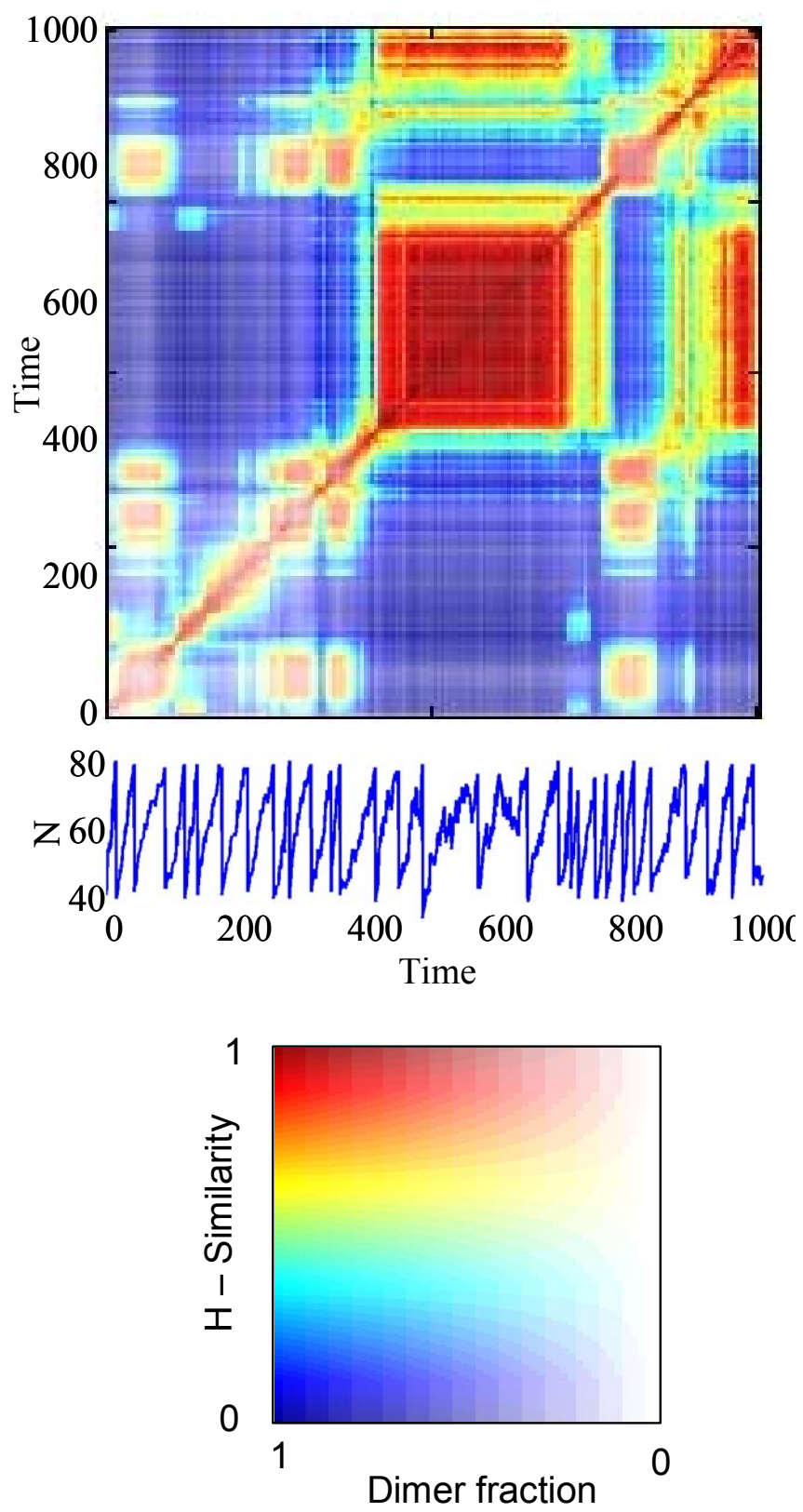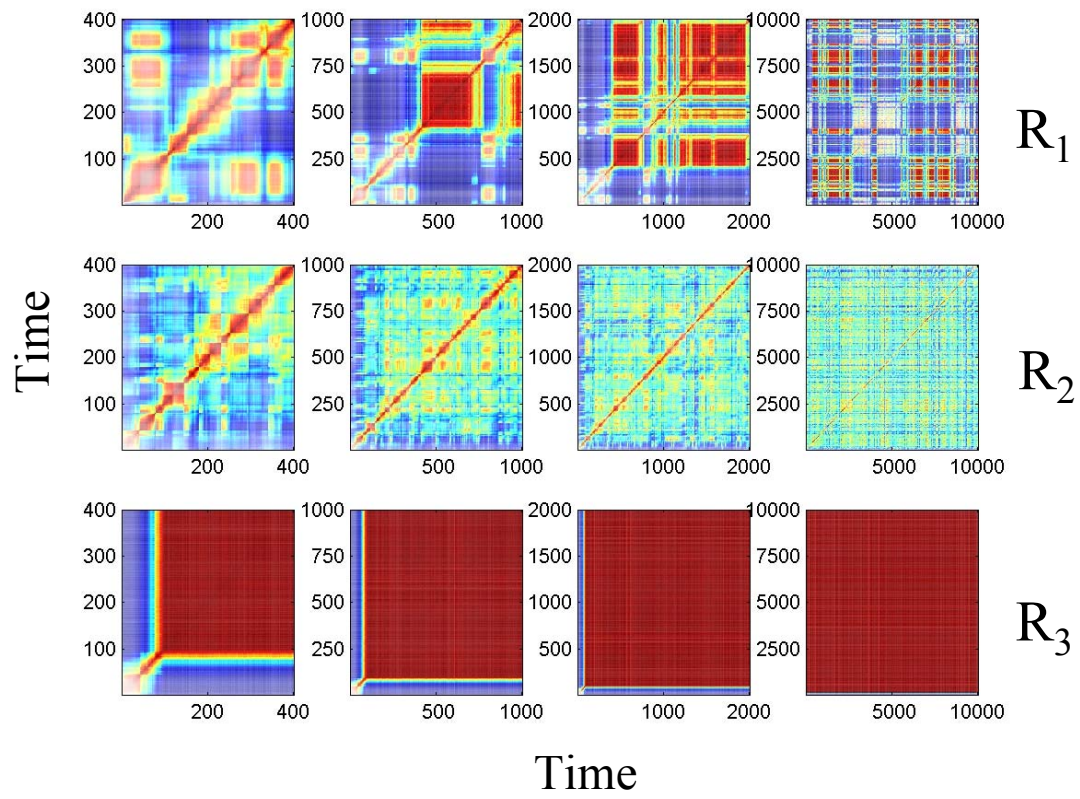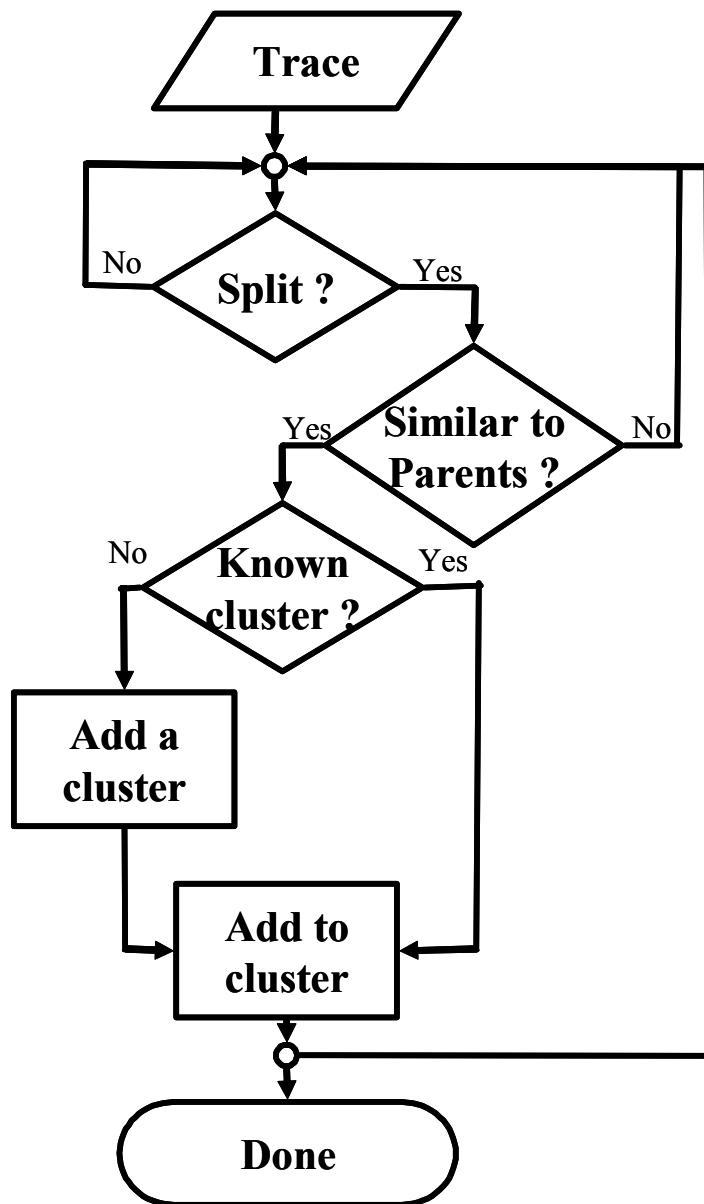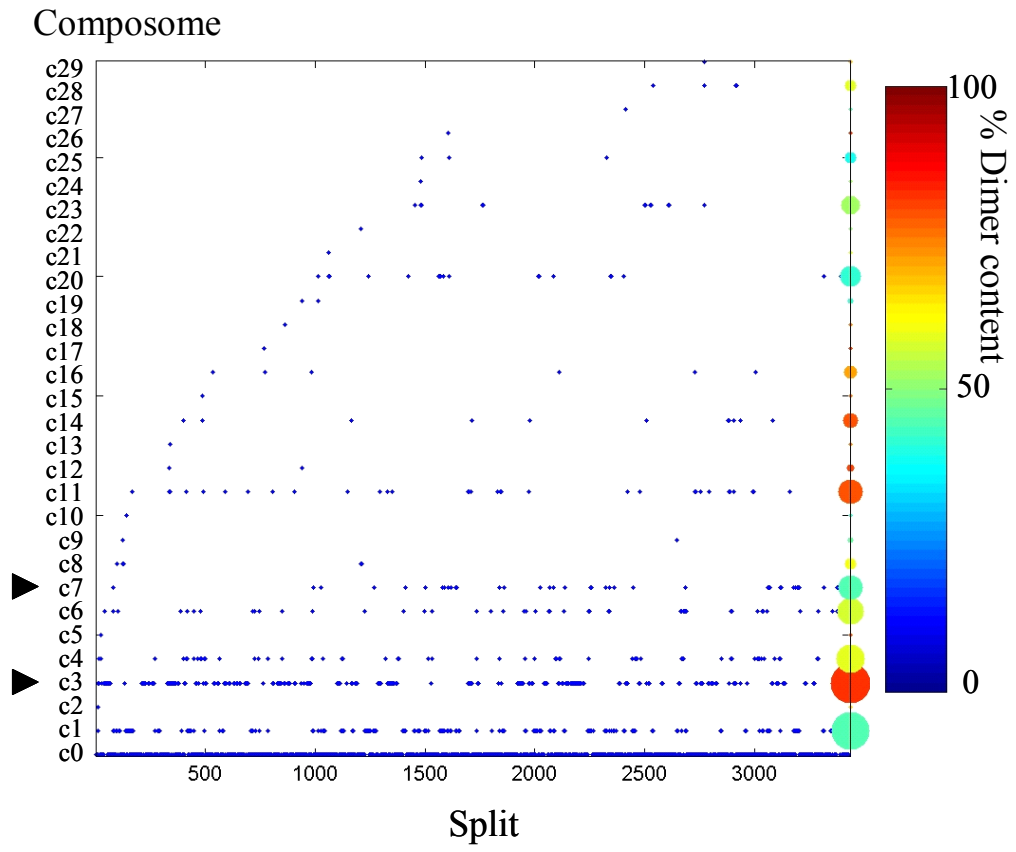
Figure 4

Figure 5A

Figure 5B

$$y = 1.419x^{0.387}$$
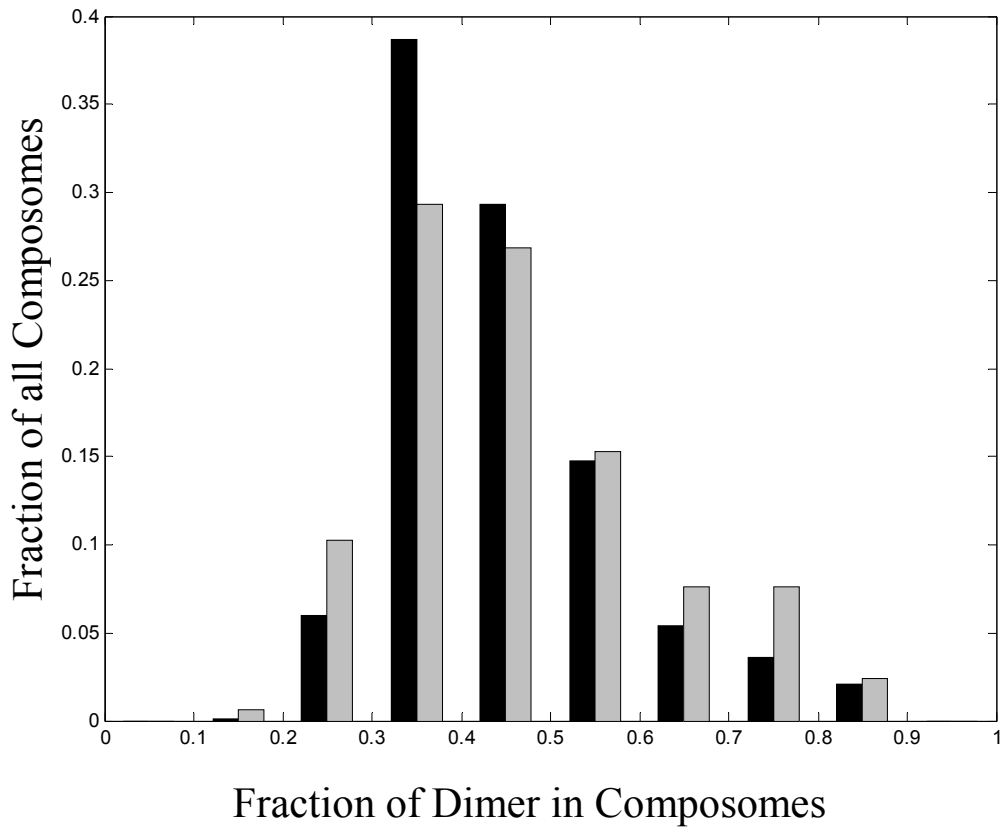
$$y = 0.323x^{0.528}$$
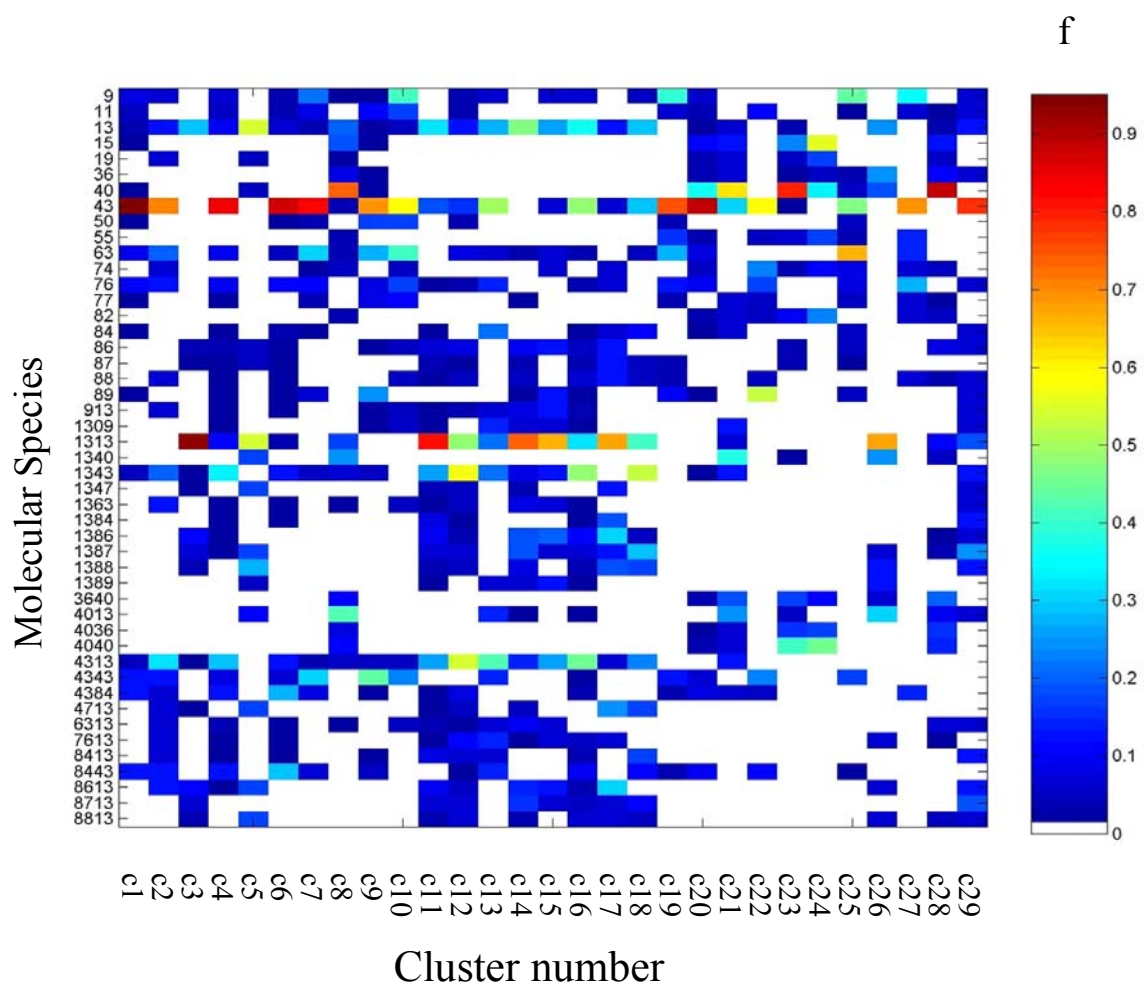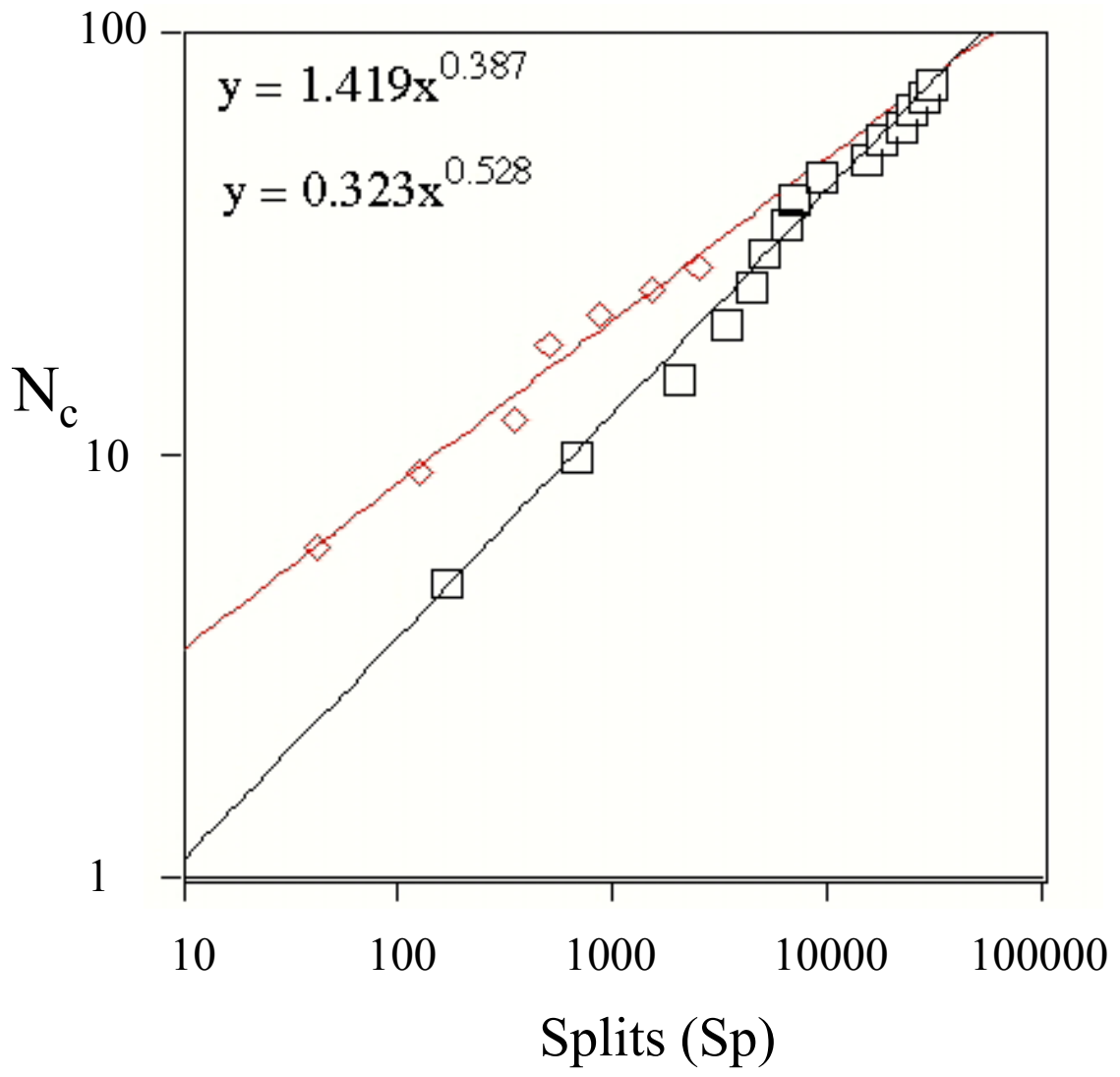
Figure 5C

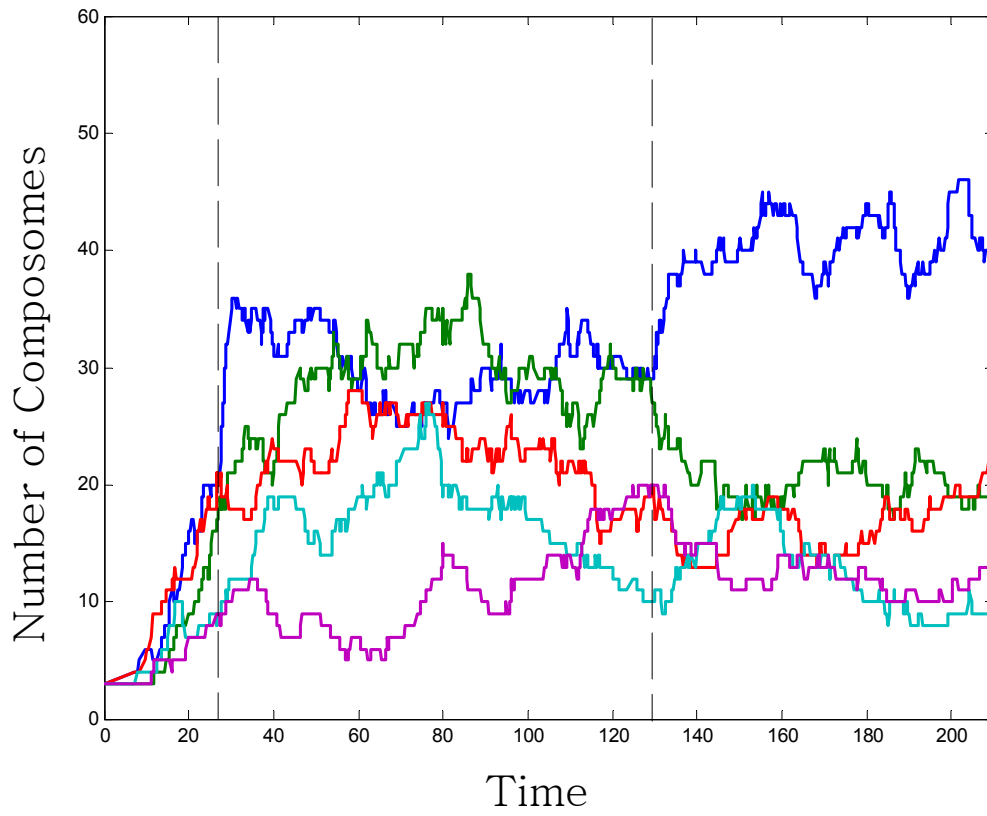C3
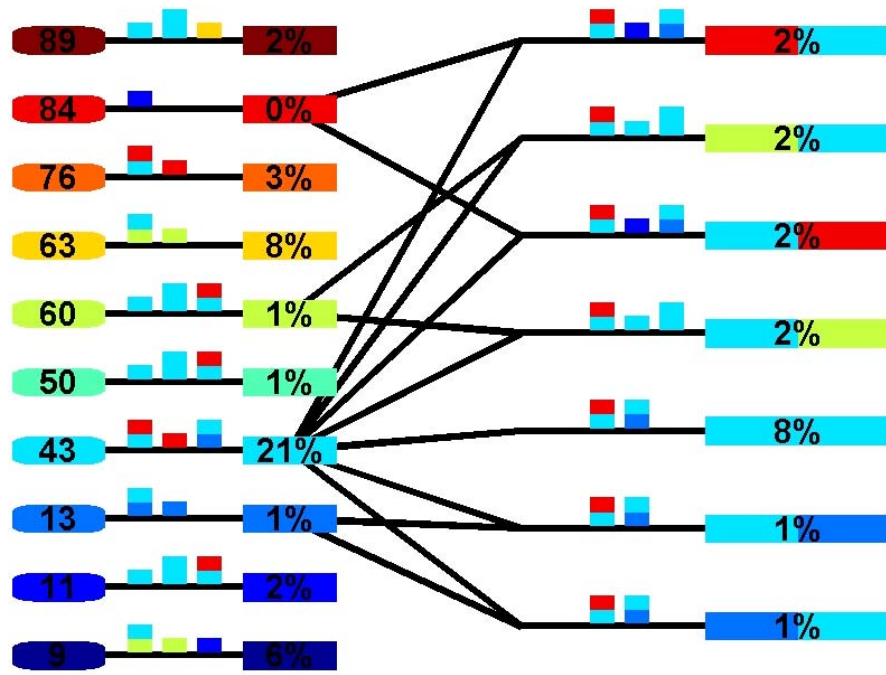
C7
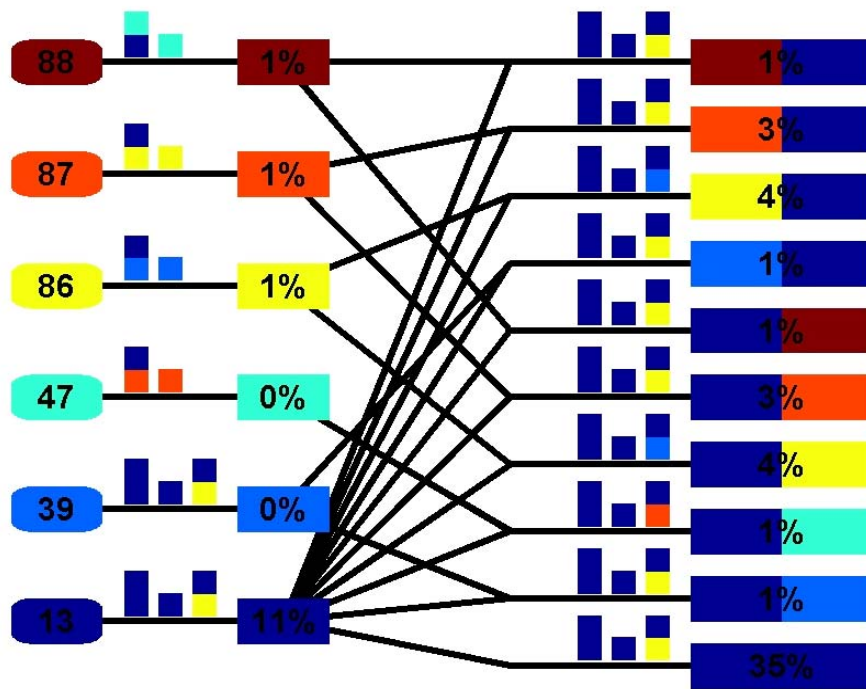
Figure 6A

C3

C7
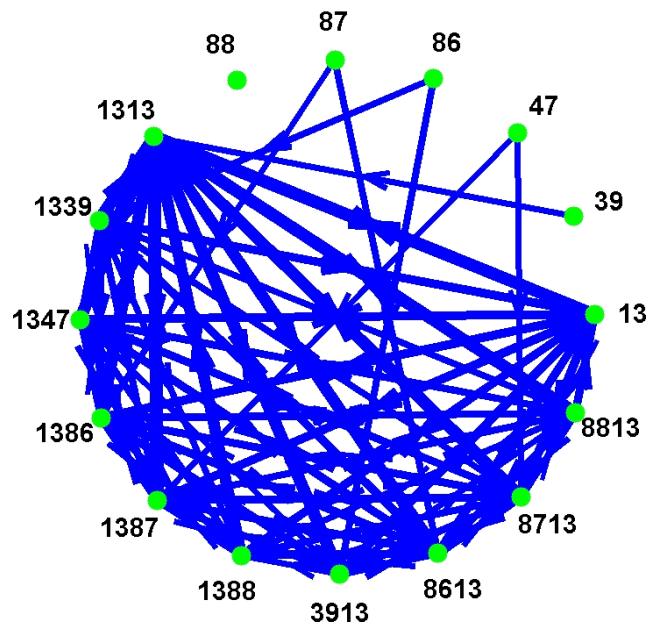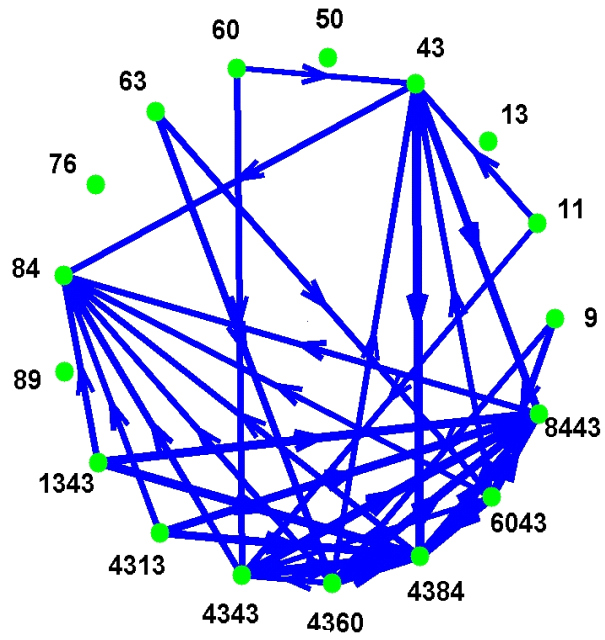
Figure 6B
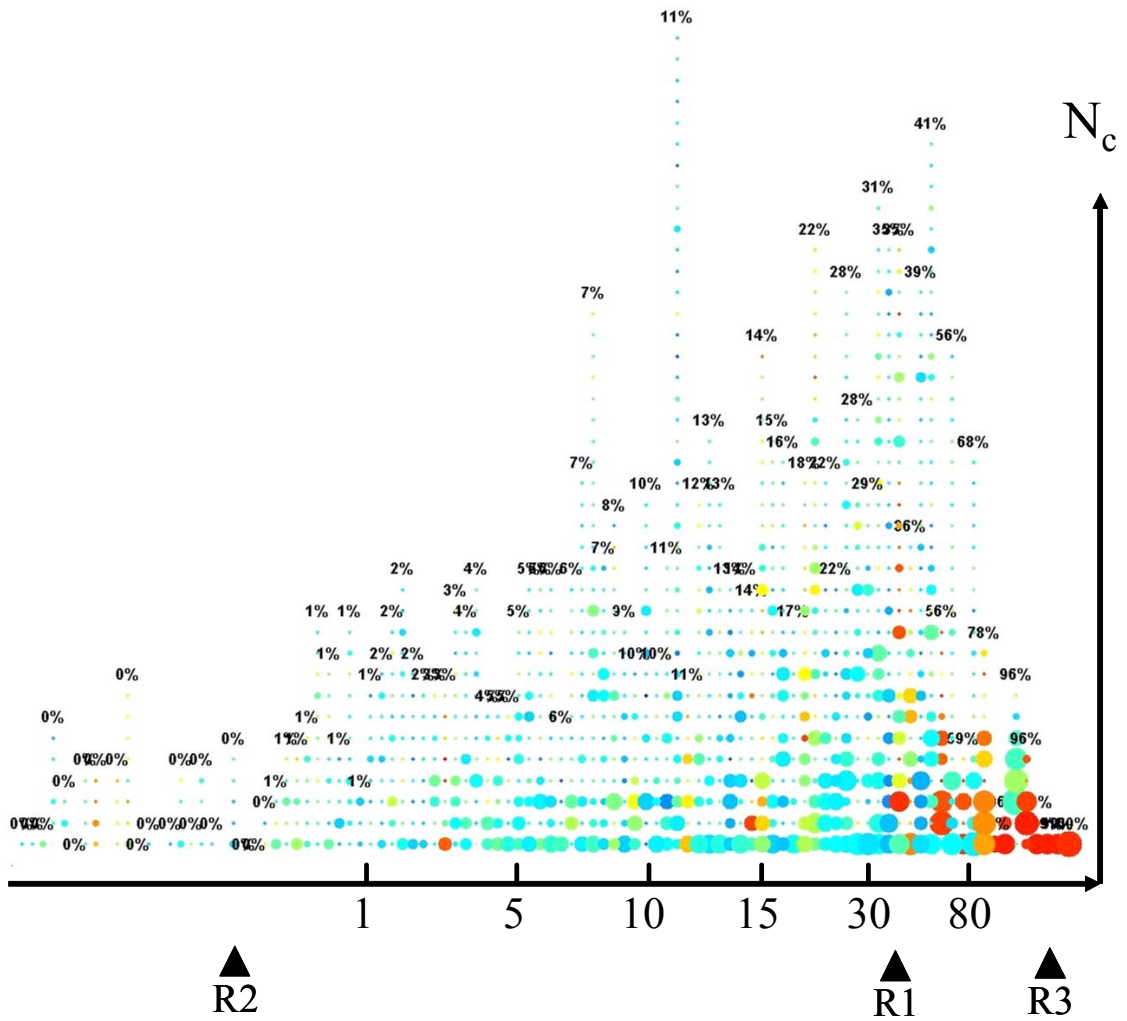
Figure 7A
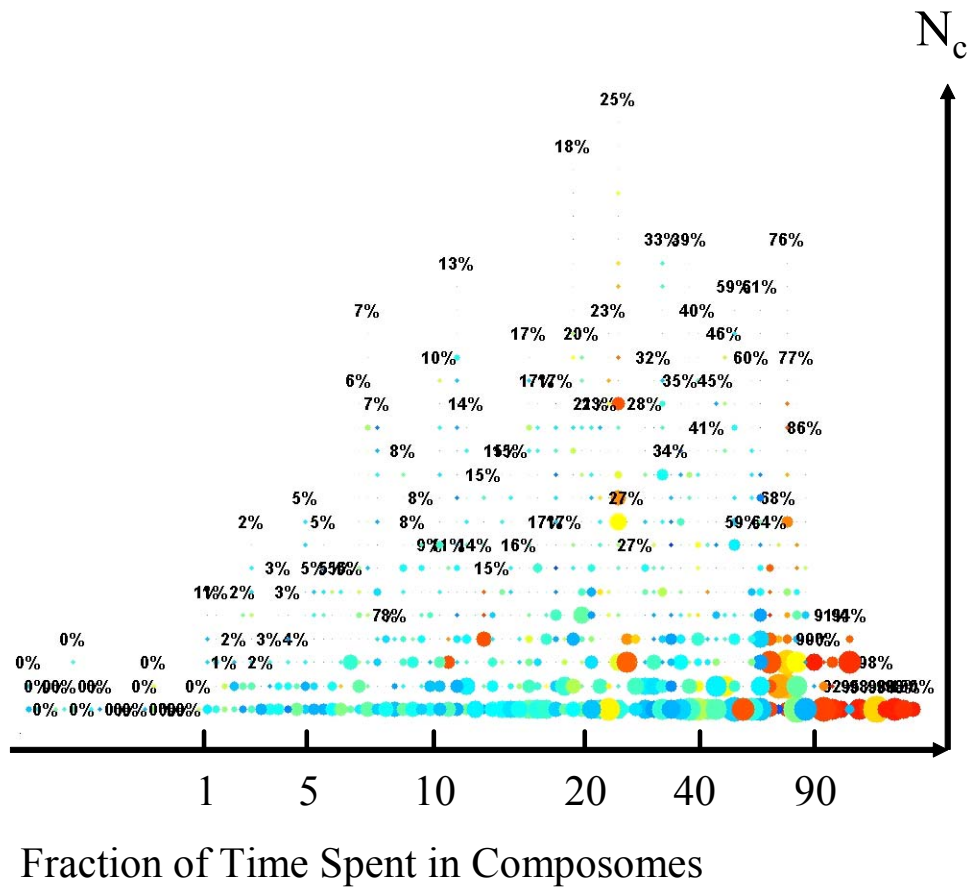
Figure 7B
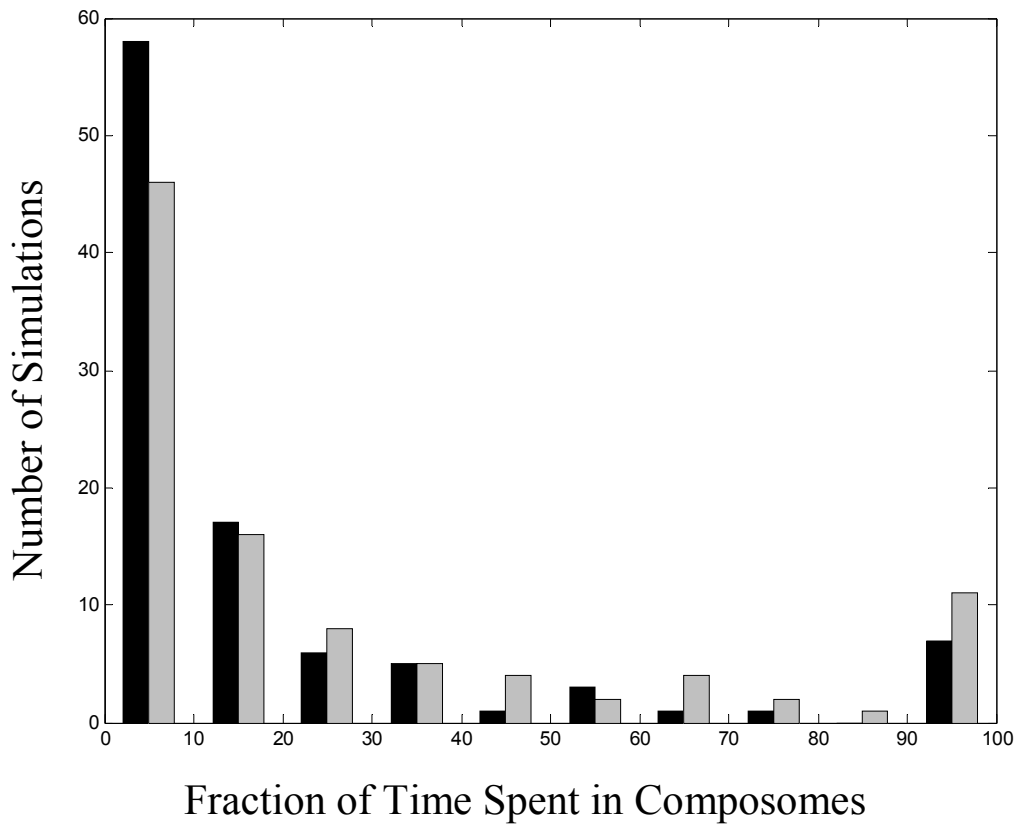
Figure 8A

Figure 8B

Figure 9A

Figure 9B