On the Potential Role of DNA in an RNA World: Pattern Generation and Information Accumulation in Replicator Systems

P. Hogeweg

Bioinformatica, University of Utrecht, Padualaan 8, 3584 NL Utrecht, Netherlands

Key Words: Computer Experiments / Nonequilibrium Phenomena / Nonlinear Phenomena

Apparently at some stage in early evolution DNA evolved to complement RNA for information storage.

In this paper we investigate under which circumstances the amount of accessible information is enhanced by using long-lived molecules for information storage.

We show that pattern formation is essential for the exploitation of information stored in long-lived molecules.

Introduction

Obviously a crucial issue in the theory of biological evolution is how information can accumulate and complexity can increase.

Complex information processing in dynamical systems requires stable configurations for information storage moving configurations for information transfer and nonlinearity (decision making) [1]. Information storage can occur as persistent activation of sub-networks (as is the case in universal cellular automata [1], hypercycles [2] or autocatalytic sets [3 – 6]. However, as is reviewed below, it appears that the information accumulation is constrained by the existence of information thresholds.

Information storage can also be realised through relatively long-lived entities, i.e. entities which survive a period in which they are not activated. The latter 'option' apparently evolved during the transition from a presumed RNA world to DNA-based living systems. Another example of the use of long-lived analogues for information storage might be the occurrence of 'memory' cells to accomplish 'immunity'.

In this paper we investigate the potential of this type of information storage assuming that the transcription (and therewith use) of this information requires activation by short-lived molecules. We show that such information is only 'retrieved' in systems in which spatial pattern formation allows for differential activation and elimination of (active) information.

Information Accumulation in Early Evolution

An important result of classical models of prebiotic evolution is the existence of information thresholds. Models which consider *non-selfreplicating* primary building blocks study the generation of autocatalytic sets; a 'percolation' threshold exists: only if the initial number of molecules be large enough to obtain percolation, an autocatalytic set forms (cf. [3 – 6]). In order to maintain systems with a large number of molecules, selfreplication should be suppressed in such systems [5].

Indeed, in the case of selfreplicative entities, information cannot be stored readily in a number of coexisting entities because of 'competitive exclusion' or 'survival of the fittest', i.e. the very properties on which the evolutionary process is based, and which leads to selective information storage within the entities themselves. However, the amount of information stored within the entities is limited by the 'error threshold' [2] and should be low in early evolution. Specific catalytic interaction structures, however, may enable the coexistence of several competing selfreplicating entities and so enhance the total amount of information which can be stored. However, in well-mixed systems, as classically studied in terms of ordinary differential equations, only cyclic interaction-structures without loose ends (hypercycles) can persist (cf. [2, 7]). Hypercycles allow only limited information accumulation, because of temporal destabilisation of longer cycles and especially because of their vulnerability to 'parasitic' mutants (i.e. mutants which are catalysed by the cycle but who are themselves not catalysts of any member of the cycle) [8].

A number of different partial solutions for the information storage problem in replicator networks have been proposed, e.g.:

- Compartmentalisation

Embedding interacting replicators in competing, replicating compartments can dynamically stabilise inherently unstable dynamical systems, e.g. by exploiting stochastic variation [9].

- Spatial pattern formation

The evolutionary dynamics of the hypercycle model is very different if it is embedded in an unstirred medium [10 – 16]. Spatial pattern formation generates new levels of selection (e.g. spirals), and therewith leads to temporal stabilisation of arbitrary long cycles and resistance to parasites. Nevertheless, unless spatial heterogeneity (gradients) are present, evolutionary dynamics leads eventually to destabilisation. Gradients enhance the dynamical persistence of hypercycles and may lead to information accumulation by partial isolation of various regions (Boerlijst and Hogeweg in prep.).

In spatially extended systems, moreover, not only are cyclic interaction structures persistent, but many more interaction topologies are feasible [16 – 19]. Nevertheless, information accumulation appears to be limited in such systems as well.

  0005-9021/94/0909-1135 $ 10.00+.25/0

- Local vs. global competition

Apart from pattern generation, spatially extended systems (as modelled in CA) differ from well mixed systems (as modelled in ODE) because of local instead of global competition [18]. Local competition in the spatial systems leads to non-localised competition in the interaction networks, and to a larger set of persistent interaction topologies, and therewith to a potential for more information storage, even if complete mixing does occur.

Information Storage and Retrieval in Cellular RNA 'Worlds'

In this paper we look at a "next stage" in evolution: i.e. the potential of long-lived molecules (*DNA*) to increase the information storage in *cells*, consisting primarily of selfreplicating entities which can mutually catalyse one another (*RNA*). *RNA* can catalyse *DNA* to produce a *RNA* analogue of itself; *DNA* is not catalytic.

We are here interested in informatic constraints for information storage and retrieval. We ignore all other constraints (such as physical/chemical). To underline this fact we will put *RNA*, *DNA* and *cell*, when they are used as name for our 'model' entities, in italics throughout this paper.

In the experiments reported here we employ all three methods of potentially increased information storage capacity discussed above: compartmentalisation, and local competition with or without spatial pattern generation. Thus, the evolutionary dynamics may 'choose' which mechanism is used (if any) to enhance information storage capacity. We investigate whether the introduction of *DNA* in such systems can enhance the amount of information 'used' (i.e. information which occurs as (selfreplicating) *RNA*), and how spatial pattern formation influences the information storage and retrieval.

To this end we define a multi-level evolutionary system, consisting of a population of cellular automata (CA's) (each CA representing one *cell*); the transition rules of the CA's model selfreplication and catalysis of *RNA*, and transcription of *DNA* (which is located in an additional layer of the CA). The CA population is subjected to (artificial) selection for 'diversity' (amount of information) i.e. number of *RNA*-species present at the time of reproduction, using a simple genetic algorithm. CA replication occurs once in 125 time steps. Gene duplication and mutation occur during CA replication. CA replication can involve 'fusion' of *cells*. This fusion takes the role of 'cross-over' in conventional genetic algorithms and simply collates two (parts of) CA's together.

The CA's are initially filled with a relatively small number of different *RNA* molecules plus just one copy of their *DNA* counterparts. In particular we used 30 species chosen randomly from the first 40 of the total (fixed) set of 160 possible *RNA* species. Accumulation of stored information occurs through gene duplication and mutation in *DNA*. 'Retrieval' of this accumulated information occurs through catalysis of *DNA* by *RNA* into its *RNA* analogue. New "*RNA*" species can also arise by "somatic mutation" (1 per time step). (For further details on this CA-based evolution system see [18, 19].)

This multilevel evolutionary system enables us to study the potential of information storage and retrieval in 'the best of all possible worlds', in the sense that we attempt to maximise information storage and retrieval by the genetic algorithm and no other constraints but those inherent in selfreplicating and competing systems are present.

Moreover, by studying a multilevel evolutionary system we hope to gain some insights on how the potential of evolutionary systems is enhanced by the mutual interaction of various levels of selection.

Experimental Design

We study matched sets of experiments, i.e. the same universe with the same initial conditions is studied with and without information storage in *DNA*, with and without "somatic mutations", and with and without pattern formation. In evolution the 'side-effects' are often more interesting and revealing than the extent to which some a-priori supplied optimisation criterion is met [20, 21]. Thus, we not only study the amount of information available to the system (i.e. the fitness criterion) during selection, but also:

- the size of the genome (including 'junk' *DNA*, i.e. parts which are not transcribed)
- how 'viable' a *cell* is in isolation (i.e. the stability of the attractor which is obtained, i.e. how dependent it is on the selection at the level of *cells*), including:
- the total size of the *RNA* pool which is expressed in isolation
- the average size of the *RNA* pool present at any one time in isolation
- the presence of permanent *RNA* species in isolation
- the dependence on the genome (if present during selection).
- the vulnerability to somatic mutations (if absent during selection)

We report in detail on one typical set of experiments. Qualitatively the same results were obtained in other sets of experiments.

Information Accumulation

Fig. 1 shows the information accumulation expressed as the average fitness (number of *RNA* species per *cell*) in the population during the selection to maximise this amount of information.

Fig. 1a shows three selection regimes in the well-mixed systems, i.e. with *DNA* only, with somatic mutations during *RNA* replication only, and with both *DNA* and somatic mutation. The presence of long-lived *DNA* appears to be necessary for survival. In absence of *DNA* the severe selec-

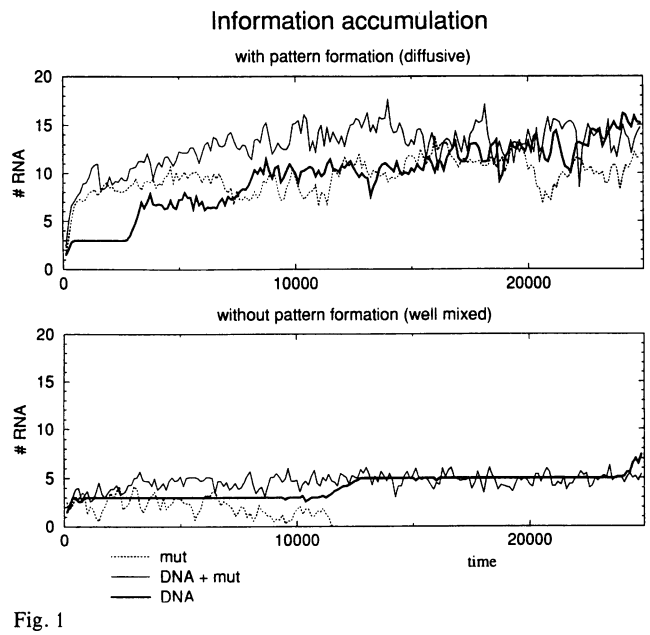

Fig. 1
Information accumulation in evolving systems, with and without pattern formation: *DNA* is neccessary for information accumulation only in absence of pattern formation.
Horizontal axis time (in *RNA* replication steps); vertical axis: average number of *RNA* species per *cell*; fat solid line: evolution with *DNA* only; thin solid line: evolution with *DNA* and somatic mutations; dotted line: evolution with somatic mutations only

tion on diversity keeps a population 'alive' during 12000 time steps, but it finally dies out. In fact at every selection step not more than half the population of *cells* is 'alive' (i.e. contains replicating *RNA*). In contrast, in the presence of *DNA* and without influx of *RNA* species through somatic mutation, the population converges on a very stable attractor of 3, then 5 and finally 8 *RNA* species per *cell*: no *cell* deaths occur and the homogeneity of the entire population obviates selection most of the time. Information accumulation occurs in 'steps' (punctuated equilibria); once in a while the appearance of a new *DNA* species through mutation or *cell* "fusion" leads to the transcription and persistent presence of *several* other *RNA*s.

Selection does play a role in maintaining the diversity (and viability) in the population with both *DNA* and somatic mutation: the population is non-homogeneous and at most selection steps some *cells* have died 'naturally'. Information content of the *cells* is much more variable: high information content, however, often leads to subsequent death.

Fig. 1b shows similar evolution regimes for the diffusive system in which pattern formation adds an extra level of selection in the multilevel evolutionary process. It is clear that spatial pattern generation significantly enhances information accumulation. However, it appears that here *DNA* is not essential for survival of the population, nor of that of the individual *cells*. Moreover, the presence of *DNA* barely increases information accumulation.

Notwithstanding the fact that *DNA* is essential in well-mixed systems and not in diffusive systems, the genome of the selected well-mixed systems is (much) smaller than that of the non-mixed pattern-generating systems (ca. 38/45, i.e. 45 genes of 38 types, in the case of the mixed system without and 70/100 with somatic mutation, whereas in the case of the pattern-generating system 85/125 genes are accumulated on average). As the gene-duplication/mutation regime is identical in all cases and amounts to an addition of ca. 3000 genes into the population (>150/cell), we conclude that there is a very strong negative selection to information accumulation in the genome of the well-mixed system, in absence of an influx of *RNA* species.

Information Retrieval

Studying the *cells* in isolation reveals differences in the role that the various levels of selection play in the maintenance of a high active information content in the *cells*.

Selection at the level of *cells* is important in absence of *DNA*, and in the mixed systems with *DNA* and somatic mutations, but much less so in all other cases. Moreover, the role of *DNA* is very different in presence or absence of pattern formation. Retrieval of the information stored in *DNA* is very limited in well-mixed systems, whereas it is extensive in pattern-generating systems, where it is, moreover, used successfully to 'defend' the system against 'harmful' invasions (by somatic mutations or otherwise).

Junk *DNA* in Well-Mixed Systems

The attractors in the mixed systems, which evolved with *DNA* and without somatic mutations, are stable and invariant, i.e. the same 8 species of *RNA* remain present in the *cells* (although complex oscillations appear in the relative amounts of the molecules). None of the *cells* dies in the 5000 steps of the simulation. Removal of the stored *DNA* does not affect the *RNA* present in the *cell* at all: the same 8 species remain. We conclude that although *DNA* is necessary for information accumulation, it renders itself entirely superfluous later on. The *RNA* attractor is not resistant to somatic mutations: adding these, with or without the presence of *DNA* kills all cells within 5000 time steps.

Thus, in well-mixed systems the most urgent need for the maintenance of diversity appears to be the exclusion of 'harmful' molecules. This is accomplished in the most diverse systems by attaining an attractor which is uninvadable by many molecules, and by severe selection at the *DNA* level such that harmful molecules are not incorporated in the genome.

We conclude that in a sense all *DNA* is 'junk' *DNA*, as it is normally never expressed. However, its presence was essential in evolution, and the *RNA* attractor and the *DNA* have co-evolved so as to render the *RNA* attractor uninvadable to the *DNA* species present in the genome. Selection at the cellular level eliminated 'harmful' *DNA*.

In contrast, in mixed systems in presence of *DNA* and somatic mutations, more *DNA* is accumulated and expressed during selection. This renders the population viable under selection at the cellular level. In isolation, the majority of the *cells* (11/18) die within 3000 time steps, but some

survive, whereas all *cells* die within 1000 time steps when the *DNA* is taken out. We conclude that the accumulation of information in the genome serves as a (be it often failing) protection against *RNA* invasions, which are not fenced of by the *RNA* attractor itself. It can do so e.g. by parasitising a parasite. However, increase of active information in these *cells* (>11), i.e. expression of much 'Junk' *DNA* most often leads to rapid extinction.

DNA for Defence in Pattern-Generating Systems

The attractors of the diffusive systems, which evolve in presence of *DNA* and without somatic mutations, are somewhat less stable than those of the mixed systems; 3 (out of 18) *cells* die in isolation (within 5000 time steps). This stability is however much larger than required in the evolving population, where average age at *cell* death is less than 1000 time steps. In contrast to the mixed systems, the attractors are very variable in time. On average there are ca. 12 *RNA* species present per *cell* at any one time, i.e. about the same as during selection and *cell* fusion in the previous 5000 time steps. Thus, selection at the *cell* level does not play a role in maintaining fitness. However, during the 5000 time steps on average, ca. 50 *RNA* species occur in each *cell*; only a few (2–5) *RNA* species are present all during this period, and some *cells* do not have any permanent *RNA* species. Thus, *RNA* species continuously are transcribed from the genome and are eliminated again. The long-lived *DNA* molecules thus serve for information preservation. Indeed, when *DNA* is removed after selection, diversity drops to ca. 6 *RNA* species. Nevertheless most *cells* remain 'viable'.

When we add somatic mutations to the isolated *cells* (with *DNA*), viability barely decreases (only 4 out of 18 *cells* die within 5000 time steps). However, these invasions of 'foreign' *RNA* increase the portion of passive information retrieved from the *DNA* pool: on average 75 different *RNA* species are present during the 5000 time step period. Average fitness remains about the same. When *cells* with *DNA* were subjected to somatic mutations during selection, they are in isolation even more resistant to invasion of 'foreign' *RNA*, but they are somewhat less diverse (ca. 10 *RNA* species).

However, if somatic mutations are present after removal of *DNA*, viability is much decreased: 14 out of 18 *cells* die within 5000 time steps, if selected without, 11 out of 18 die if selected with somatic mutations; the others maintain only limited diversity (7–8 RNA species) notwithstanding the influx. Similarly when selection is performed without information storage in *DNA*, and with somatic mutations the resulting attractors have also limited viability in isolation: here 12 out of 18 *cells* died within 5000 time steps.

We conclude that in the presence of pattern formation, information storage in *DNA* is essential for information preservation to allow increase of active information (here fitness), and most importantly to serve as an effective defence system against invasions.

Discussion and Conclusions

We have studied the role of long-lived entities, *DNA*, in information accumulation and retrieval in evolving replicator systems (*RNA* worlds). By using a multilevel evolutionary process, we could focus on interesting features in this complex system.

We have found different strategies for the maintenance of high diversity dependent on the presence or absence of pattern formation in the system. The difference is most pronounced in absence of somatic mutations. In presence of somatic mutations a compromise between the two strategies is employed. In both strategies the role of *DNA* is very different.

Patterns in the *RNA* pool enable differential invasion and elimination of information into the system, and therefore *DNA* can play an active role in the preservation of information, both by supplying information and by eliminating 'harmful' information. By this mechanism the *cells* become 'excitable' and 'adaptable' – properties which we often associate with 'living' systems. Here it is obtained simply as a side-effect of selection for information accumulation and pattern generation.

Contrariwise the only 'strategy' available for well-mixed systems for maintaining a relatively large amount of information is to limit the invadability of the attractor attained. This renders DNA, which played an important role in the accumulation of this information, functionless ('junk'), although selection at the *cell* level limits the type of 'junk' admitted into the *DNA* pool. Nevertheless *DNA* may also here defend the system to a limited extent to invasions; expression of too much *DNA* leads, however, to extinction.

One may object that our systems 'do not do' anything with the information which they accumulate and retrieve. This is indeed the case. However I think that maintenance of information may be the limiting factor; information which can be preserved will, because of its preservation, be 'selected' as substrate for the maintenance of other information, and thus obtain, secondarily a function. Because biotic systems do not have any a-priori goals it can simply do what it is able to do.

References

[1] S. Wolfram, Universality and complexity in cellular automata, Physica 10D, 15 (1984).
[2] M. Eigen and P. Schuster, The hypercycle: A principle of selforganization. Springer, Berlin 1979.
[3] S.A. Kauffman, Autocatalytic sets of proteins, J. Theor. Biol. *119*, 1 (1986).
[4] S.A. Kauffman, The origins of order; selforganization and selection in evolution, Oxford University Press 1993.
[5] W. Fontana, Algorithmic Chemistry, in: Artificial Life II, pp. 159–209, ed. by C.G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, Addison Wesley 1991.
[6] R.J. Bagley and J.D. Farmer, Spontaneous emergence of a metabolism, in: Artificial Life II, pp. 93–140, ed. by C.G. Langton, C. Taylor, J.D. Farmer and S. Rasmussen, Addison Wesley 1992.
[7] J. Hofbauer and K. Sigmund, The theory of evolution and dynamical systems, Cambridge University Press 1988.

[8] J. Maynard Smith, Hypercycles and the origin of life, Nature *180*, 445 (1979).
[9] E. Szathmary and L. Demeter, Group selection of early replicators and the origin of life, J. Theor. Biol. *128*, 463 (1987).
[10] M.C. Boerlijst and P. Hogeweg, Spiral wave structure in prebiotic evolution: Hypercycles stable against parasites, Physica D *48*, 17 (1991).
[11] M.C. Boerlijst and P. Hogeweg, Selfstructuring and Selection, in: Artificial Life II, pp. 159–209, ed. by C.G. Langon, C. Taylor, J.D. Farmer and S. Rasmussen. Addison Wesley 1992.
[12] R. May, Hypercycles spring to life, Nature *353*, 607 (1991).
[13] C. Streissler, Autocatalytic networks under diffusion, Doctoral thesis, Universität Wien 1992.
[14] P. Chacon, J.C. Nuno and F. Moran, An. Quim. *6*, 379 (1993).
[15] R. Hecht, P. Schuster and C. Streissler, Replicator networks with intermediates. Paster, Jena 1994.
[16] M.C. Boerlijst and P. Hogeweg, Evolutionary consequences of spatial pattern formation, in: S.C. Muller and T. Plesser (eds.) Spatio-temporal organization in non-equilibrium systems, pp. 26–28, contributions to the Dortmunder Dynamische Systemen Woche, June 1992, Project Verlag 1992.
[17] M. Nowak and R.M. May, The spatial dilemmas of evolution. Int. J. Bifurcation Chaos *3*, 35 (1993).
[18] P. Hogeweg, Multilevel evolution: replicators and the evolution of diversity. Physica D, in press 1994.
[19] P. Hogeweg, Multilevel molecular evolution: Pattern generation and its evolutionary consequences, in: Chaos and Complexity. Editions Frontieres, Paris 1994.
[20] P. Hogeweg and B. Hesper, Evolution as pattern processing: TODO as a substrate for evolution, in: J.A. Meyer and S.A. Wilson (eds.) From Animals to Animats, MIT press 1991.
[21] P. Hogeweg and B. Hesper, Evolutionary dynamics and the coding structure of sequences: Multiple coding as a consequence of high mutation rates, Comput. Chem. *16*, 171 (1992).

Presented at the Joint Meeting and International Bunsen Discussion Meeting of the Deutsche Bunsen-Gesellschaft für Physikalische Chemie "Self-Organization of Biopolymers" in Jena, April 10th to 13th, 1994 E 8704