Hyperstructures

Vic Norris[1], Patrick Amar[2], Pascal Ballet[3], Gilles Bernot[2], Franck Delaplace[2], Maurice Demarty[1], Jean-Louis Giavitto[4], Camille Ripoll[1], Michel Thellier[1] and Abdallah Zemirline[3]

[1]*Laboratoire des Processus Intégratifs Cellulaires, UPRESA CNRS 6037, Faculté des Sciences & Techniques , Université de Rouen, 76821, Mont-Saint-Aignan, France*
[2]*Laboratoire de Méthodes Informatiques, CNRS UMR 8042, Université d'Evry, 91025 Evry cedex, France*
[3]*Laboratoire des Interfaces Machines Intelligente, Université de Bretagne Occidentale, Brest, France*
[4]*Laboratoire de Recherche en Informatique, Université Paris-Sud, Orsay, France*

Contents

Abstract

New concepts may prove necessary to profit from the avalanche of sequence data on the genome, transcriptome and proteome and to relate this information to cell physiology. Here, we focus on the concept of hyperstructures in which a variety of types of molecules are brought together to perform a function. The processes responsible for hyperstructure formation include changes in enzyme affinities due to metabolite-induction, transertion, and elevated local concentrations of proteins and their binding sites on DNA and RNA. We review the evidence for the existence of hyperstructures responsible for the initiation of DNA replication, the sequestration of newly replicated origins and for cell division. We interpret cell cycle progression in terms of hyperstructure dynamics. Finally, we speculate on how a variety of *in silico* approaches could be combined to develop new concepts in the form of an *Integrated* or *Imaginary* cell – *I-cell* – which would undergo selection for growth and survival in a world of artificial microbiology.

1 Introduction

Molecular biology and biochemistry have provided a wealth of information about how RNA polymerases transcribe DNA into RNA and how ribosomes then translate mRNA into proteins, about the nature of those proteins and lipids and form membranes, and about other important molecules. Model organisms such as the bacterium *Escherichia coli* are invaluable in making sense of this information. The 4.6 Mb genome of *E. coli* has been sequenced (Blattner *et al*., 1997) and was found, at the time, to have 4288 protein-coding genes (cf 5885 in the eukaryote *Saccharomyces cerevisiae*) of which 38% had no attributed function. But even when all genes are ascribed functions, how are we to interpret this information and use it to predict phenotypes? The challenge is to understand how cells organise their myriad constituents and processes. To explain how the concept of hyperstructures may help us, here, we briefly review the bacterial cell cycle, focussing on the problem of division, and then discuss hyperstructures. We do this in the light of different questions: Why might a hyperstructure language be useful? What are hyperstructures? How do they form? How do they interact? How might they guide cells through state space to control growth, adaptation, differentiation and the cell cycle? We then discuss how the hyperstructure concept may help in exploiting the information provided by genome sequencing and how it may be tested. Finally, we advocate the construction of an integrated cell, I-cell, as a new approach to the study of biological complexity.

2 The bacterial cell cycle

The principal events in the bacterial cell cycle include:

- *Initiation of chromosome replication from a single origin of replication*
- *The sequestration of newly replicated origins of replication*
- *Chromosome separation*
- *Chromosome segregation*
- *Cell division*
- *Inactivation of the division site*

In the case of cell division, it is still not clear how this event is timed, positioned and coupled to other events. The earliest known protein to act in *E. coli* is the tubulin-like FtsZ which migrates from the cytoplasm to a mid-cell location on the membrane where it assembles into a ring-like structure and where it recruits other division proteins. What lies upstream of FtsZ? Is it yet

another protein or is it something else? We have shown that FtsZ can interact directly with phospholipid membranes in the absence of other proteins (Alexandre *et al* ., 2001). This is consistent with a major role for membrane dynamics in the regulation of the cell cycle as is the finding that membrane domains around the chromosomes differ from the domain at the future site of division (Fishov & Woldringh, 1999), this latter presumably being related to the large domains of cardiolipin observed at the division sites and poles (Mileykovskaya & Dowhan, 2000).

The constraints on a solution to the division problem for *E. coli* are that the division site must be:

• *in the right place – midcell – to give daughters of similar sizes*
• *between chromosomes to avoid producing a DNA-less cell*
• *formed at the right time in the cycle – perhaps to give the right DNA/mass ratio?*
• *formed at the right rate – to avoid, for example, cells getting bigger and bigger*
• *of the right nature – to allow membranes to curve and fuse whilst controlling ion and lipid fluxes*

It is in the context of trying to find a solution to this problem that we present hyperstructures.

3. Why invoke a hyperstructure language?

Cells survive and sometimes grow by somehow orchestrating millions of molecules of thousands of types to adapt to the environment and to proceed through the cell cycle. This entails cells solving the combinatorial problem of negotiating the immensity of state space since if each gene in *E. coli* were in either an on (transcribed) or an off (untranscribed) state, there would be 2^{4000} or 10^{1200} on-off patterns of gene expression (Kauffman, 1996). But there is more than this, there is also the epigenetic trap – cells in a population should not all have the same phenotype (else, for example, a single catastrophe would be more likely to wipe them all out). Exploring state space effectively boils down to: How can cells be both efficient and robust? We argue that the answer is that cells rely on an intermediate level of organisation – *hyperstructures.*

4. What are hyperstructures?

There are two sorts of hyperstructures:

• Non-equilibrium hyperstructures are large structures of diverse molecules – genes, mRNAs, proteins, ions, lipids – that depend on a flow of energy/material for their existence. These hyperstructures are assembled to serve a specific function and are disassembled when no longer functional (Figure 1).
• Equilibrium or quasi-equilibrium hyperstructures are large structures that are not dependent on a flow of energy for their existence. They are not assembled or disassembled according to whether they are required to serve a function. Our discussion of equilibrium hyperstructures will be very limited here.

Examples of possible non-equilibrium hyperstructures include:

An initiation hyperstructure responsible for starting the initiation of replication of the chromosome. This hyperstructure contains the DnaA protein and certain of the sites on DNA to which it binds (Norris *et al*., 2001). DnaA is the key protein in initiation in *E. coli* and binds to 9mer TTA/TTNCACA and 6mer AGATCT sites present in the origin of replication and in

certain replication-related genes (Speck *et al*., 1999). DnaA polymerisation (Weigel *et al*., 1999) and a fluid, acidic, membrane domain (Castuma *et al*., 1993; Fralick & Lark, 1973) are required for DnaA to be active in initiation.

A replication hyperstructure comprising the protein SeqA plus the key enzymes in DNA replication along with the genes that encode them (Norris *et al*., 2000). SeqA sequesters newly replicated origins of replication and prevents them from being used more than once within a substantial portion of the cell cycle (hence preventing a flurry of initiations when a single initiation signal is given). SeqA is found in clusters (Onogi *et al*., 1999). It binds to GATC sequences and it polymerises.
A cell division hyperstructure comprising the 10 or so division proteins (including FtsZ) plus enzymes involved in peptidoglycan synthesis together with the genes that encode them, many located together in the *dcw* cluster at the 2 min position on the chromosome (Buddelmeijer *et al*., 1998; Norris & Fishov, 2001).

Other hyperstructures include a DNA compaction hyperstructure possibly involving MukB which can form foci (Ohsumi *et al*., 2001); a nucleolus-like hyperstructure responsible for ribosome synthesis and assembly (Lewis *et al*., 2000; Woldringh & Nanninga, 1985); a chemotaxis hyperstructure comprising chemotactic receptors such as Tsr with the kinase CheA and the transducing protein CheA (Bray *et al*., 1998; Stock & Levit, 2000) plus, we propose, the genes encoding these proteins. Factors in the formation of possible hyperstructures for transport and glycolysis (Mitchell, 1996; Norris *et al*., 1999; Velot *et al*., 1997) are discussed below.
Examples of possible equilibrium hyperstructures:

These include immiscible domains within the condensed chromosome. This immiscibility occurs in the context of an ordered liquid, with the DNA closely layered by a regular twist (Bouligand & Norris, 2001), a situation that may minimize entangling and facilitate co-expression of the genes within a domain. Our discussion of equilibrium hyperstructures will be very limited here.

5. How do non-equilibrium hyperstructures form?

There are several complementary possibilities:

5.1 Metabolite-induction

The idea is that:

• Non-equilibrium hyperstructures form when the cell is actively engaged in processing substrates and disappear when they are not
• These hyperstructures include enzymes in the same pathway and their genes
• Formation of certain of these hyperstructures may involve an interplay between diffusion in 2-D and 3-D in the sense that enzymes confined to domains in the 2-D membrane interact with other enzymes or groups of enzymes diffusing in the 3-D cytoplasm (Figure 1)

The evidence consistent with this scenario has been advanced for the existence of metabolons which are assemblies of the enzymes that act in succession in a pathway (Velot et al., 1997). Of course, such metabolons may themselves associate into larger hyperstructures. In the case of secretion, substrate binding promotes assembly of the 3 components of the ABC exporters of Gram negative bacteria e.g. in *Erwinia chrysanthemi* the substrate (protease) binds to PrtD (an ABC protein) which then binds to PrtE (membrane fusion protein) and which binds to PrtF (outer

membrane protein) (Letoffe *et al*., 1996). In the case of glycolysis, the glycolytic pathway can be extracted as an equimolar complex of 1.65 megaDa that reveals compartmentation of substrates (Mowbray & Moses, 1976). In the case of import, sugar-specific phosphotransferase system permeases consist of EIIC and EIID in the membrane and EIIA and EIIB in the cytoplasm; EIIA is phosphorylated by HPr in a reaction catalysed by EI with phosphate from phosphoenolpyruvate; E2s+E1+HPr probably form a complex (Norris et al., 1999). The idea here is that successive enzymes in the same pathway can be activated to bind to one another in a *vertical* organisation. A complementary idea is that a single species of enzyme can be activated to oligomerize by substrate (Torshin, 1999); indeed, the full enzymatic activity of glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase and enolase – all glycolytic enzymes – results from their association. Again, this *horizontal* organisation could help nucleate and stabilise hyperstructures (Figure 2).

5.2 Local concentrations

The phenomenon of oligomeric proteins binding to specific sites on DNA has been invoked to explain the operation of the *lac* and lambda repressors (Revet *et al*., 1999). It might also be invoked to explain the sequestering of newly replicated origins of replication by the protein SeqA (Onogi et al., 1999). There are variations of this theme with, for example, the possibility that proteins such as the histone-like protein HU, which binds to both RNA and DNA (Balandina *et al*., 2001), could play important roles (see below and Figure 3).

5.3 Transertion

Transertion is the coupled transcription, translation and insertion into and through membranes of proteins. The cytoplasmic membrane is composed of a wide variety of lipids and proteins so, if these proteins have lipid affinities, small proteolipid domains form. High rates of transertion may create a critical density of inserted nascent proteins that is sufficient for small proteolipid domains to fuse into large ones and so nucleate hyperstructure assembly (Norris, 1995) (Figure 4). For example, it might be supposed that the high density of transertion of the ATP synthetase components, which have lipid affinities (Arechaga *et al*., 2000; Ksenzenko & Brusilow, 1993), should result in assembly of an ATP synthesis hyperstructure.

5.4 Translated mRNA is protected from RNases and enzymes in metabolons are protected from proteases

It has been suggested that enzymes in complexes are more likely to escape the attention of proteases than when those enzymes are not in complexes (Miller, 1996). An extension of this idea is that the partitioning of enzymes into a hyperstructure protects them from proteases (providing the latter are excluded from the hyperstructure). Hence an enzyme which has been assembled into a hyperstructure *because* of its activity is thereby preserved (i.e. active enzymes are preferentially protected). A similar argument is that mRNA translated within a hyperstructure could be preferentially protected from RNases on the outside of the hyperstructure.

6. How do hyperstructures interact?

6.1 Shared lipid affinities creates shared membrane domains

It can be argued that proteins with lipid preferences may congregate with those lipids in a positive feedback fashion to form the membrane domain part of a hyperstructure (see *5.3 Transertion*). Similarly, it might be expected that hyperstructures characterised by enrichment for a particular lipid would also tend to associate.

6.2 Shared binding proteins create shared cytoplasmic compartments

The idea is that certain abundant proteins may participate in the assembly of several different types of hyperstructures. This would enable a synergy whereby the progressive formation of a group of hyperstructures responsible for a set of functions would aid the recruitment of other related hyperstructures fulfilling complementary functions. Candidates for these proteins include the DNA-binding proteins IHF, FIS, and HU (for references see (Ussery *et al*., 2001)). IHF can modulate the transcriptional activity of promoters by influencing the looping of upstream DNA; the consensus site of IHF binding, YAACTTNTTGATTTW, lies within many repetitive extragenic palindromic sequences. FIS binding to upstream regions can enhance the transcription of highly expressed genes; the consensus for the FIS binding site is weak with estimates of its numbers ranging from 6 to 68000. HU binds to DNA with no evident sequence preference and, in so doing, influences the interaction of regulatory proteins with their specific sites on the DNA (Bonnefoy & Rouviere-Yaniv, 1992); HU also recognizes certain specific structures of both DNA and RNA with very high affinity and, for example, binds to the mRNA for RpoS (Balandina et al., 2001; Kamashev & Rouviere-Yaniv, 2000). In addition, there are over a 100 known activators and repressors of transcription in *E. coli* (Ouzounis *et al*., 1996) and it may be expected that these will control the synthesis of certain oligomeric proteins important in the assembly of different – but complementary – hyperstructures.

6.3 Shared codon preferences

There are strong compositional asymmetries in codon and amino acid usage depending on the orientation of the genes with respect to DNA replication and on the nature of the proteins encoded. This has led to predictions of different compartments for the syntheses of different proteins (Danchin & Henaut, 1997).

6.4 Water preferences

Water exists as species with different structures and chemical properties that affect the distribution and activity of cellular constituents (Robinson *et al*., 1999; Wiggins, 1990). An important but difficult question is the extent to which the water preferences of the constituents of hyperstructures might determine hyperstructure formation and interaction.

6.5 Oscillations/vibrations

The Min system, which is involved in the selection or inactivation of the division site, oscillates with a periodicity of around 1 minute in *E. coli* (Raskin & de Boer, 1999). There are numerous oscillatory processes in eukaryotes of which the oscillation of protons and of NAD(P)H in neutrophils is particularly exciting (Petty & Kindzelskii, 2001). Such oscillations are candidates for playing global as opposed to local organising roles. Relating them to the dynamics of hyperstructures is a problem that has still to be addressed.

7. Cell division

The regulation of cell division can now be considered in terms of the dynamics of hyperstructures. It has been argued that one of the functions of the bacterial cell cycle is to generate daughter cells with different phenotypes since this would allow the population to both explore all the possibilities for growth offered by the environment and be ready for a sudden catastrophic change (Norris et al., 2001; Segre *et al*., 2000). In this scenario, during the run-up to

initiation, the mass to DNA ratio increases and certain hyperstructures become 'stronger' by attracting ever more of the cell's resources (such as the transcriptional and translational apparatus) whilst other hyperstructures are weakened and disappear (Norris et al., 2001)(Figure 5). This results in a drop in the diversity of hyperstructures, some of which release DnaA as they dissociate, a DnaA-initiation hyperstructure forms, and replication of the chromosome begins. Now suppose that short FtsZ polymers are associated with glycolytic and other hyperstructures so that FtsZ is effectively sequestered (noting that, at least in eukaryotes, tubulin is associated with glycolytic enzymes (Lloyd & Hardin, 1999)). This leads us to consider two possibilities. One is that the FtsZ-sequestering hyperstructures are temporarily disrupted by chromosome replication to release FtsZ which can then participate in division. The other, complementary, possibility is that the changing activity of the phosphotransferase system/glycolytic hyperstructure directly leads to its own disassembly (for example, its capacity might exceed demand and lead to feedback inhibition) and releases FtsZ. This would be consistent with the advance in divisions in synchronous cultures of *E. coli* induced by addition of the non-metabolisable, glucose analogue α-methylglucoside (Fishov, 1994) and the delay in divisions induced by transfer to a rich growth medium (Kepes & Kepes, 1985).

Before trying to put it all together, we should bear in mind that, all else being equal, the rates of transcription of two copies of the same gene diverge if this gene is –vely regulated *in trans* but +vely *in cis* (Norris & Madsen, 1995). The –ve regulation *in trans* could result from a repressor diffusing through the cytoplasm to each separate stretch of DNA whilst the +ve regulation *in cis* could result from an RNA polymerase transcribing a gene making this particular stretch of DNA accessible to another polymerase. This leads to the conclusion, surprising for many microbiologists, that two identical chromosomes in the same cytoplasm (which contain many such genes) *therefore* have different patterns of gene expression. The same argument can be made in terms of hyperstructures: a set of genes is expressed from one chromosome to form part of a hyperstructure; this assembly involves positive feedback between the constituent ions, lipids, proteins and nucleic acids since as the density of one constituent in a region increases, the probability increases that the density of another constituent will also increase. In the context of a cell in which hyperstructures compete for existence (that is, negative regulation *in trans*), the result is a highly structured, asymmetric cell in which each future daughter cell has a different set of hyperstructures associated with it and these sets differ in their composition of lipids, ions, water structures, proteins, mRNA and expressed genes (Figure 6). At present, it is difficult to discriminate between the different ways in which the principal proteolipid domains around the chromosomes could create a division site (red, dotted arrows in Figure 6). In one scenario, the site would simply consist of the *interface* between the two domains whilst in the other scenario, the site would consist of a distinct *domain* between the principle two domains. There are, of course, permutations of these possibilities. The essence of our proposal is that hyperstructure dynamics could achieve:

- separation of chromosomes during replication
- differentiation of both chromosomes and membrane
- the right place for a site to attract and activate division enzymes (between the chromosomes)
- the right time for the creation of a division site (after chromosome segregation)
- the right nature for a division site - a potential non-bilayer
- coupling between replication, segregation and cell division
- a calcium flux (down the concentration gradient)
- orchestration of membrane-activated kinases, proteases *etc.*

8. The advantage of organisation at the level of hyperstructures

It has been observed that the difficulty of administering a laboratory is proportional to the square of the number of members of the laboratory, N^2 (Bok, 1983). This difficulty, D, is reduced if the individuals are put into N_1 groups such that D equals the square of the number of groups (to reflect group interactions) plus the square of the number of individuals in each group N_0^2 (to reflect interactions within groups) times the number of groups:

$D = N_1^2+(N_0^2)N_1$

Hence $D = (N/N_0)^2+(N_0^2)N/N_0$

And $D = N^2/N_0^2+N_0N$

To minimise D,

• $D/\bullet\ N_0 = -2N^2/N_0^3+N$

Hence the difficulty is at a minimum when

$N_0 = (2N)^{1/3}$

This formula helps to give us a feel for the numbers of hyperstructures that may exist in a cell. Just considering proteins, for example, a bacterium containing of the order of a million interacting proteins would be expected to have around a hundred hyperstructures. The existence of this intermediate level of organisation therefore means that the problem of generating a limited number of coherent phenotypes that are adapted to survival and/or growth is greatly simplified. Navigation through the immensity of state space becomes a choice between 100 or so hyperstructures rather than 4000 plus genes – 2^{100} on-off combinations rather than 2^{4000}. To generate a coherent phenotype, for example, enzymes appropriate for growth in cold oxygenated conditions should not be synthesized in the same cell at the same time as those for growth in hot anaerobic conditions. Coherence can be achieved because cells can manage the relatively few common factors required to bring together a particular set of hyperstructures. The existence of hyperstructures also allows, we speculate, bacterial cells to regulate DNA replication and cell division so as to create heterogeneous populations that can both grow and survive unexpected challenges.

9. Using the hyperstructure concept to exploit sequence data

Of the numerous *in silico* approaches possible, we focus here on cellular automata which are used to model many physical and biological phenomena (Vichniac, 1984). Once the units that constitute the automata have been assigned initial states, the evolution of these states can then depend on both the previous history of the state and on the state of neighboring units. Hence, cellular automata can be particularly suitable for modeling the dynamics of interactions between molecules in 3 dimensions. We now use cellular automata to illustrate how they might be used to model the effects on hyperstructure assembly of the following:

9.1 Metabolite-induction

To determine the values of the parameters governing the formation of hyperstructures in bacteria, we have constructed a preliminary version of a cellular automaton program (with features of multi-agent systems) that simulates the dynamics of the localization of the PTS and glycolytic

enzymes in both a 2 dimensional membrane and a 3 dimensional cytoplasm (Le Sceller *et al*., 2000). Each unit volume represents a 10nm*10nm*10nm cube in a cell that can have a maximum volume of 200*200*200 unit volumes or 8μm^3. This is more than sufficient to represent *E. coli* which in certain growth conditions has a volume of 2 cubic microns. Each cubic unit volume in the membrane is surrounded by 8 other unit volumes and each unit volume in the cytoplasm is surrounded by 26 others. At each time step, all enzymes are considered in a random order. Each can move into a free neighboring unit volume. In this preliminary study, there was a structuring of both membrane and adjacent cytoplasm and hyperstructures were generated containing up to 500 enzymes.

9.2 Transertion

To model the anchoring effect of transertion on nascent proteins (Figure 4), a proportion of the PTS Enzymes II (for example) could be permanently confined *in silico* to a patch of the membrane. An important parameter may therefore be the *area* over which these proteins are inserted. It is not easy to obtain this area experimentally with current techniques (but see the NanoSIMS below). However, this may be an instance when the simulation reveals whether hyperstructure formation is very sensitive to the area of transertion and therefore whether energy should be invested in performing the relevant experiments.

9.3 Lipid preferences

The cosegregation of proteins with the lipids for which they have pronounced affinities is a potent way to produce domains. This process may be simulated in the 'membrane' of cellular automata given these affinities. Below (*10.3*), we suggest a series of experiments that could lead to consensus sequences for lipid binding and hence a way, ultimately, to convert sequence information into the 'lipidome' and facilitate the simulation of the distribution of all membrane proteins.

9.4 Local concentrations

Using cellular automata to model local concentrations might exploit knowledge of DNA-binding proteins and their sites providing DNA can also be introduced into the model. One way to achieve this would be to divide the chromosome into chunks comparable in size to proteins. Each chunk would be constrained in its diffusion by a function inversely proportional to the distance between the chunk in question and another chunk. It may also prove necessary to make efforts to model reptation, the constrained movement of polymers in a crowded solution.

9.5 DNA distribution

DNA curvature, flexibility and stability have been analysed for 18 fully sequenced bacterial genomes (Pedersen *et al*., 2000). This reveals many significant structural features including a set of 20 regions with identical and extreme structural properties that are proposed to function as topological domain boundaries. These features are presumably related to the properties of proteins such as HU (see *6.2*) which binds preferentially to unusual structures such as kinked or cruciform DNA (Bonnefoy *et al*., 1994; Kamashev & Rouviere-Yaniv, 2000). The challenge is to translate this information into a dynamic 3-D model taking into account that much of the DNA is probably in a cholesteric form. One model that might be tested *via* cellular automata (as in *9.4*) is that HU both binds to these curved regions and self-associates such that curved regions are stacked at the edges of twisted liquid crystalline regions. In such a model, the terminus region,

which has high curvature, low flexibility and low helix stability (Pedersen et al., 2000), might be expected to exhibit a distinctive packing.

9.6 Parallel approaches

In an activity-based vision of the cell, only a subset of its constituents is important in determining the phenotype of the cell at any one time (Norris, 1998). This subset comprises those constituents that are *active* where *active* is considered to mean being transcribed for a gene, being translated for a mRNA, and catalysing a reaction for an enzyme. Belonging to this active subset requires a competition between constituents that were active in the previous time period (the *status quo* factor) and constituents that act in synergy with one another (the coherence factor).

In this section, we describe a new implementation of cellular automata or units based on the related idea that only a few unit volumes are potentially active, that is either contain a molecule or a have a neighbouring unit containing a molecule. The advantage is that memory is not needed to store these empty units. This leads to a time and memory efficient approach for computing the successive generations of the units. The overall state of the system is determined by the content of all the units at a given time. Computing the next generation means determining the new state of the system after the application of all the local rules to each unit. This process must not depend on the order the units are examined and, ideally, each unit is treated independently of all the other units. The standard way to represent the 3-D space is to use a array of structures to address each unit that often contains only a number. Using this method, it is easy to determine the neighbourhood of a unit by a simple transformation of its coordinates, and then access the array to get the values of the neighbouring units. The major drawback is that we must store *all* the units, even the empty ones.

In our approach, we also represent the space by a three coordinate system, but we store in the computer memory *only* the active or potentially active units (i.e. those that are filled or next to filled units). This reduces the memory cost and allows us either to reduce the size of each unit to have a more accurate simulation, or to simulate a larger space.

The potentially active cells are stored in a hash table which allows a very fast access time, comparable to the access time of a 3-D array, if a good hash function is used along with an adapted strategy to resolve collisions. This low cost implementation of the state of the system can be used to reduce the time used to compute each generation if an extra cost is paid by duplicating the representation of the space: the local rules are applied to each active unit using the values from the first, *current* space and the result is stored in the second, *new* space. After all the units of the first space have been processed and the second space is complete, the second space becomes the current space and the next generation can be computed.

Since the current space is only accessed for *reading* values whilst the new space is only accessed for *writing* results, the current space can be freely accessed by multiple processes without synchronisation. The new space can be split into parts that can be computed separately on a multiprocessor with a consequent dramatic reduction in computation time. Each process requires its own part of the current space but also acts on a surrounding layer of single units in the parts treated by other processes. Since each process only accesses the part of another process at the boundary, each part can be stored locally in a multi-computer networked environment.

In the 12 by 12, 2-D example (Figure 7), process **P1** only needs to access the first 7 lines (0 to 6) of the current space to compute the first 6 lines of the new space, while process **P2** needs to

access the last 7 lines (5 to 11) of the current space to compute the other half of the new space. Since there is no read/write conflict between **P1** and **P2** no synchronization is needed. This is another advantage of the inherent parallelism of this implementation.

9.7 Hyperstructure movements and reactions

Interactions *between* hyperstructures are proposed to result in a pre-divisional cell with one set of hyperstructures in one half the cell and a different set in the other half. Such sets of hyperstructures may be formed on the basis of common lipids, ions, binding proteins and/or water properties. Movements of hyperstructures are nicely illustrated by the SeqA-replication hyperstructure that, during the cell cycle, goes from a single focus to two foci that then migrate to the one-quarter and three-quarter positions (Ohsumi et al., 2001; Onogi et al., 1999). To model how interactions between hyperstructures might lead to redistribution of hyperstructures within the cell, we consider a cellular automaton model in which several hyperstructures can be represented simultaneously in a coarse-grained way (so that the units are bigger than single macromolecules).

The idea presented in this section entails providing local rules to reproduce molecular reaction and diffusion using cellular automata. The difference between our approach and typical reaction-diffusion processes is that the molecule concentration (in a specific position) is boolean: *true* if there is a set of molecules, *false* if there are none. One of the simplest systems has only one type of molecule on a 2-D grid (environment). Focusing on a particular molecule and its neighbourhood, it is clear that a unit plus the 8 adjacent units is a square of side 3 units. If we suppose that the molecule can move or stay in the same place, the molecule will have 9 possible positions (Figure 8). With cellular automata, the state of a unit depends on its neighbourhood. Thus, a local rule must be used to determine whether a unit in the 2D-grid becomes *true* (has molecules) or *false* (is empty). The idea is to invert the arrow direction in the previous figure. Thus, if an empty cell is adjacent to one filled cell, it has a probability of $1/9^{th}$ to become filled. We can apply this to any neighborhood.

Given a unit in a 2D-grid, the probability of the unit to become *true* (filled) is $p = n / 9$ where n is the number of filled units into its neighbourhood. More generally, for a *dim*-dimensional environment (a grid of dimension *dim*), the probability of one cell to become *true* is:

$$p = n / N \qquad (1)$$

where $N = 3^{dim}$. Figure 9 shows the results of this local probabilistic rule (rule 1) with a *60x60* grid at three different times.

To construct a multi-molecule hyperstructure, the next stage consists in putting together different types of molecules. In this case, the value of a unit is not a boolean but an integer included between *0* and *nb* (number of types of molecule). In this way, rule (1) becomes

$$\begin{aligned} p_0 &= n_0 / N \\ p_1 &= n_1 / N \\ p_2 &= n_2 / N \\ &\ldots \\ p_{nb} &= n_{nb} / N \end{aligned} \qquad (2)$$

where n_i is the number of molecules of type *i* into the neighborhood and $N = 3^{dim}$ into the *dim*-dimension grid. An empty cell is of type *0*.

To choose the future type of a cell among all the possibilities, we consider a real random number *A* bounded by *0* (included) and *1* (excluded). The decision rules are the following:

$$\begin{array}{c} \text{if } A \in [0, p_0[\text{ then the considered cell is of type } 0 \\ \text{if } A \in [p_0, p_0+p_1[\text{ then the considered cell is of type } 1 \\ \ldots \\ \text{if } A \in [p_0+p_1+\ldots+p_{nb-1}, p_0+p_1+\ldots+p_{nb}[\text{ then the considered cell is of type } nb. \end{array} \quad (3)$$

Figure 10 shows the results of this local probabilistic rules with a *60x60* grid at three different times and for two types of molecule.

To study hyperstructures containing many different molecules that can perform chemical reactions, we allow two different molecules to react to produce another type of molecule:

$$mol_i + mol_j \rightarrow mol_k \quad (4)$$

To do this, we add the following simple rule (5) to rule (3):

if the *considered cell* is of type *i* (resp. *j*) and the type chosen thanks to rule 3 is *j* (resp. *i*) then the cell will have the type *k* (according to rule 4) (5)

Figure 11 shows the evolution of the system in a *60x60* grid at three different times, for two types of molecule and that react together according to rules (4) and (5) to produce a third type of molecule.

To observe the formation of and interaction between hyperstructures, we introduce the notion of affinity between molecules. In our example (Figure 12), molecules of type 1 are activated and can therefore bind one another. Molecules of type 2 and 3 have similar behaviours. Moreover, molecules of type 1 can react with molecules of type 2 to produce molecules of type 3. Figure 12 shows fluxes of molecules leading to the formation of a hyperstructure:

- Molecules of type 1 come from the top of the cellular automata and bind together
- Molecules of type 2 come from the bottom of the cellular automata and bind together too
- Molecules of type 1 react with molecules of type 2 to produce molecules of type 3
- Molecules of type 3 bind together and with molecules of type 1 and type 2.

10 Experimental aspects

10.1 NanoSIMS

Visualizing hyperstructures directly with conventional techniques has been difficult since it requires the co-localization of such disparate elements as proteins, mRNA, genes and lipids at the 50 nm scale. In secondary ion mass spectrometry, a section of biological material is subjected to a beam of ions that pulverizes it to release secondary ions that are filtered by mass spectrometry to allow an image to be obtained [Thellier, 1993 #1084]. Recent developments in NanoSIMS technology are very promising since the new generation of machines provides resolution at the

scale required and allows detection of isotopically marked probes to proteins and nucleic acids. This opens up the exciting possibility of studying hyperstructures by imaging simultaneously both nucleic acids and up to 10 different proteins at a resolution intermediate between light and electron microscopy.

10.2 Optical waveguide lightmode spectroscopy (OWLS)

In the case of glycolysis, we lack details of the exact abundance of proteins such as phosphoglucose isomerase**,** fructose -1,6-P2 aldolase, triose-P isomerase, glyceraldehyde 3-phosphate dehydrogenase A complex, and phosphoglycerate kinase. Although we can obtain these *via* radioactive labeling and 2-D gel electrophoresis, there are attractive, recent techniques such as those based on isotope-coded affinity tags (Gypi *et al*., 1999). More seriously, we lack details of the constants of affinity of the PTS and glycolytic enzymes. These could be obtained using OWLS in experiments with purified proteins and substrates (Ramsden, 1993). By introducing and removing the substrates, it may also prove possible in these experiments to estimate the period of time for which an enzyme remains active (i.e. has a higher affinity constant) once its substrate has gone (Ricard *et al*., 1998).

10.3 MALDI-MS and ES-MS

We are presently using sensitive techniques of mass spectrometry to explore the possibility that concomitant with overproduction of a membrane protein is a compensatory overproduction of the lipid for which it has an affinity (Arechaga et al., 2000). If this approach is successful, a semi-automated, general strategy might be developed in which bacteria are transformed with plasmids each containing a different peptide (from a random library); the idea is to obtain thousands of colonies, each containing lipids resulting from the overproduction of a particular peptide. Mass spectrometry and sequencing would then match lipids and peptides. The data would be used to try to derive consensus sequences to be used to interpret the genome and construct a 'lipidome'.

10.4 Atomic Force Microscopy (AFM) and the Langmuir-Blodgett technique

Langmuir-Blodgett monolayers of phospholipids, which assemble at the air-water interface, followed by transfer to a solid support and inspection with AFM, provide a powerful combination of techniques for studying FtsZ interaction with membranes and may constitute the beginnings of an *in vitro* division system (Alexandre et al., 2001). The characteristics of the lipids used along with the values of parameters obtained for factors that interact with FtsZ, such as calcium, GTP and other division proteins, might be used to try to construct an *in silico* model of the division process.

11. I-cell

Developing new concepts may prove essential to a full understanding of how a cell works. To test and develop such concepts, we advocate the construction of an *Integrated* or *Imaginary* cell – *I-cell* – which would undergo selection for growth and survival in a world of artificial chemistry (Dittrich & Banzhaf, 1998). The unit volumes that constitute an I-cell would be inspected at each time step and, according to the molecule(s) found, the appropriate entry would be consulted in a table containing a large number of 'biological' functions (Norris & Le Sceller, 2001). These functions would determine the interactions of the molecule with its neighbours and also, via global functions, with distant molecules. The I-cell would be fed according to different regimes and, depending on the functions implemented, would grow and eventually divide; I-cells

would be analysed after selection over several generations. Combinatorial problems would be reduced if an activity-based vision of the cell were adopted in which only a subset of constituents would be consulted at each time step; this subset would correspond to constituents that play an active role in coherent cell states via a mechanism based in part on global functions and termed competitive coherence (Norris, 1998). An I-cell might, for example, offer a way to discover the importance of a particular organising process, for example, one based on water structure or tensegrity. An I-cell might even be used to see whether new laws of complexity emerge as the number of organising processes in the system increases.

Acknowledgements

We thank Genopole and the Conseil Regional de l'Ile de France for support.

References

Alexandre, S., Cole, G., Coutard, S., *et al*. (2001). Interaction of FtsZ protein with a DPPE film. Colloids and Interfaces B in press.

Arechaga, I., Miroux, B., Karrasch, S., *et al*. (2000). Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale overproduction of the b subunit of F_1F_0 ATP synthase. FEBS Letters 482: 215-219.

Balandina, A., Claret, L., Hengge-Aronis, R., *et al*. (2001). The *Escherichia coli* histone-like protein HU regulates *rpoS* translation. Molecular Microbiology 39: 1069-1079.

Blattner, F.R., Plunkett III, G., Bloch, C.A., *et al*. (1997). The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453--1462.

Bok, J. (1983). Un modèle d'auto-organisation : le principe de moindre difficulté. In L'auto-organisation. De la physique au politique. (Dumouchel, P. and Dupuy, J.-P., eds.). Paris. Seuil,

Bonnefoy, E. and Rouviere-Yaniv, J. (1992). HU, the major histone-like protein of *E. coli*, modulates the binding of IHF to *oriC*. EMBO Journal 11: 4489-4496.

Bonnefoy, E., Takahashi, M. Rouviere-Yaniv, J. (1994). DNA-binding parameters of the HU protein of *Escherichia coli* to cruciform DNA. Journal of Molecular Biology 242: 116-129.

Bouligand, Y. and Norris, V. (2001). Chromosome separation and segregation in dinoflagellates and bacteria may depend on liquid crystalline states. Biochimie 83: 187-192.

Bray, D., Levin, M.D. Morton-Firth, C.L. (1998). Receptor clustering as a cellular mechanism to control sensitivity. Nature 393: 85-88.

Buddelmeijer, N., Aarsman, M.E.G., Kolk, A.H.J., *et al*. (1998). Localisation of cell division protein FtsQ by immunofluorescence microscopy in dividing and non-dividing cells of *Escherichia coli*. Journal of Bacteriology 180: 6107-6116.

Castuma, C.E., Crooke, E. Kornberg, A. (1993). Fluid membranes with acidic domains activate DnaA, the initiator protein of replication in *Escherichia coli*. Journal of Biological Chemistry 268: 24665-24668.

Danchin, A. and Henaut, A. (1997). The map of the cell is in the chromosome. Current Opinion in Genetics and Development 7: 852-854.

Dittrich, P. and Banzhaf, W. (1998). Self-evolution in a constructive binary string system. Artificial Life 4: 203-220.

Fishov, I. (1994). Do oscillations control the bacterial cell cycle? In What is controlling Life? (Gnaiger, E., Gellerich, F.N. and Wyss, M., eds.), pp. 221-225. Innsbruck, Austria. Innsbruck University Press,

Fishov, I. and Woldringh, C. (1999). Visualization of membrane domains in *Escherichia coli*. Molecular Microbiology 32: 1166-1172.

Fralick, J.A. and Lark, K.G. (1973). Evidence for the involvement of unsaturated fatty acids in the initiation of chromosome replication in *Escherichia coli*. Journal of Molecular Biology 80: 459-475.

Gypi, S.P., Rist, B., Gerber, S.A., *et al*. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 17: 994-999.

Kamashev, D. and Rouviere-Yaniv, J. (2000). The histone-like protein HU binds specifically to DNA recombination and repair intermediates. EMBO Journal 19: 6527-6535.

Kauffman, S. (1996). At home in the Universe, the search for the laws of complexity., Penguin, London.

Kepes, F. and Kepes, A. (1985). Postponement of cell division by nutritional shift-up in *Escherichia coli*. Journal of General Microbiology 131: 677-685.

Ksenzenko, S.M. and Brusilow, W.S.A. (1993). Protein-lipid interactions of the proteolipid c subunit of the *Escherichia coli* proton-translocating adenosinetriphosphatase. Archives of Biochemistry and Biophysics 305: 78-83.

Le Sceller, L., Ripoll, C., Demarty, M., *et al*. (2000). Modelling bacterial hyperstructures with cellular automata. Interjournal Paper 366: http://www.interjournal.org.
Letoffe, S., Delepelaire, P. Wandersman, C. (1996). Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein mediated exporters is ordered and promoted by substrate binding. EMBO Journal 15: 5804-5811.
Lewis, P.J., Thaker, S.D. Errington, J. (2000). Compartmentalization of transcription and translation in *Bacillus subtilis*. EMBO Journal 19: 710-718.
Lloyd, P.G. and Hardin, C.D. (1999). Role of microtubules in the regulation of metabolism in isolated cerebral microvessels. American Journal of Physiology 277: C1250-1262.
Mileykovskaya, E. and Dowhan, W. (2000). Visualization of phospholipid domains in *Escherichia coli* by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange. Journal of Bacteriology 182: 1172-1175.
Miller, C.G. (1996). Protein degradation and proteolytic modification. In Escherichia coli and Salmonella. (Neidhardt, F.C., ed.), Vol. I, pp. 938-954. Washington D.C. American Society for Microbiology,
Mitchell, C.G. (1996). Identification of a multienzyme complex of the tricarboxylic acid cycle enzymes containing citrate synthase isoenzymes from *Pseudomonas aeruginosa*. Biochemical Journal 313: 769-774.
Mowbray, J. and Moses, V. (1976). The tentative identification in *Escherichia coli* of a multi-enzyme complex with glycolytic activity. European Journal of Biochemistry 66: 25-36.
Norris, V. (1995). Hypothesis: transcriptional sensing and membrane domain formation initiate chromosome replication in *Escherichia coli*. Molecular Microbiology 15: 985-987.
Norris, V. (1998). Modelling *E. coli*: the concept of competitive coherence. Comptes Rendus de l'Academie des Sciences 321: 777-787.
Norris, V., Demarty, M., Raine, D., *et al*. (2001). Hypothesis: hyperstructures regulate initiation in *Escherichia coli* and other bacteria. Biochimie : in press.
Norris, V. and Fishov, I. (2001). Hypothesis: membrane domains and hyperstructures control bacterial division. Biochimie 83: 91-98.
Norris, V., Fralick, J. Danchin, A. (2000). A SeqA hyperstructure and its interactions direct the replication and sequestration of DNA. Molecular Microbiology 37: 696-702.
Norris, V., Gascuel, P., Guespin-Michel, J., *et al*. (1999). Metabolite-induced metabolons: the activation of transporter-enzyme complexes by substrate binding. Molecular Microbiology 31: 1592-1595.
Norris, V. and Le Sceller, L. (2001). A cellular automaton approach to modelling a bacterium. International Conference of Systemics, Cybernetics and Informatics : in press.
Norris, V. and Madsen, M.S. (1995). Autocatalytic gene expression occurs *via* transertion and membrane domain formation and underlies differentiation in bacteria: a model. Journal of Molecular Biology 253: 739-748.
Ohsumi, K., Yamazoe, M. Hiraga, S. (2001). Different localization of SeqA-bound nascent DNA clusters and MukF-MukE-MukB complex in *Escherichia coli* cells. Molecular Microbiology 40: 835-845.
Onogi, T., Niki, H., Yamazoe, M., *et al*. (1999). The assembly and migration of SeqA-Gfp fusion in living cells of *Escherichia coli*. Molecular Microbiology 31: 1775-1782.
Ouzounis, C., Casari, G., Valencia, A., *et al*. (1996). Novelties from the complete genome of *Mycoplasma genitalium*. Molecular Microbiology 20: 898-900.
Pedersen, A.G., Jensen, L.J., Brunak, S., *et al*. (2000). A DNA structural atlas for *Escherichia coli*. Journal of Molecular Biology 299: 907-930.
Petty, H.R. and Kindzelskii, A.L. (2001). Dissipative metabolic patterns respond during neutrophil transmembrane signaling. Proceedings of the National Academy of Science U.S.A. 98: 3145-3149.

Ramsden, J.J. (1993). Review of new experimental methods for investigating random sequential adsorption. J. Statist. Phys. 73: 853-877.
Raskin, D.M. and de Boer, P.A. (1999). Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. Proceedings of the National Academy of Science U.S.A. 96: 4971-4976.
Revet, B., von Wilcken-Bergmann, B., Bessert, H., *et al*. (1999). Four dimers of λ repressor bound to two suitably spaced pairs of λ operators form octamers and DNA loops over large distances. Current Biology 9: 151-154.
Ricard, J., Gontero, B., Avilan, L., *et al*. (1998). Enzymes and the supramolecular organization of the living cell. Information transfer within supramolecular edifices and imprinting effects. Cellular and Molecular Life Sciences 54: 1231-1248.
Robinson, G.W., Cho, C.H. Urquidi, J. (1999). Isobiestic points in liquid water: Further strong evidence for the two-state mixture model. Journal of Chemical Physics 111: 698-702.
Segre, D., Ben-Eli, D. Lancet, D. (2000). Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proceedings of the National Academy of Science U.S.A. 97: 4112-4117.
Speck, C., Weigel, C. Messer, W. (1999). ATP- and ADP-DnaA protein, a molecular switch in gene regulation. EMBO Journal 18: 6169-6176.
Stock, J. and Levit, M. (2000). Signal transduction: hair brains in bacterial chemotaxis. Current Biology 10.: R11-4.
Torshin, I. (1999). Activating oligomerization as intermediate level of signal transduction: analysis of protein-protein contacts and active sites in several glycolytic enzymes. Front. Biosci. 4: D557-570.
Ussery, D., Larsen, T.S., Wilkes, K.T., *et al*. (2001). Genome organisation and chromatin structure in *Escherichia coli*. Biochimie 83: 201-212.
Velot, C., Mixon, M.B., Teige, M., *et al*. (1997). Model of a quinary structure between Krebs TCA cycle enzymes: a model for the metabolon. Biochemistry 36: 14271-14276.
Vichniac, G.Y. (1984). Simulating physics with cellular automata. Physica D 10: 96-116.
Weigel, C., Schmidt, A., Seitz, H., *et al*. (1999). The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. Molecular Microbiology 34: 53-66.
Wiggins, P.M. (1990). Role of water in some biological processes. Microbiological Reviews 54: 432-449.
Woldringh, C.L. and Nanninga, N. (1985). Structure of the nucleoid and cytoplasm in the intact cell. In Molecular Cytology of *Escherichia coli*. (Nanninga, N., ed.), pp. 161-197. London. Academic Press.

Figure 1

Formation of a non-equilibrium hyperstructure due to changes in the affinity of its constituent enzymes for one another. Enzymes E1 can only diffuse in the plane of the membrane whilst the other enzymes, E2 to E7 diffuse in the cytoplasm. The binding of a substrate, such as a sugar, to the E1 enzymes leads to an increase their affinity for one another and their assembly into an E1 domain. On binding its substrate, each enzyme in the pathway acquires an increased affinity for the following enzyme. This results in the assembly of metabolons E1 to E7 and the assembly of the hyperstructure (here, a group of metabolons). Note that transcription of the genes encoding E1 to E7 and the simultaneous translation of the mRNA may help the assembly of the hyperstructure.

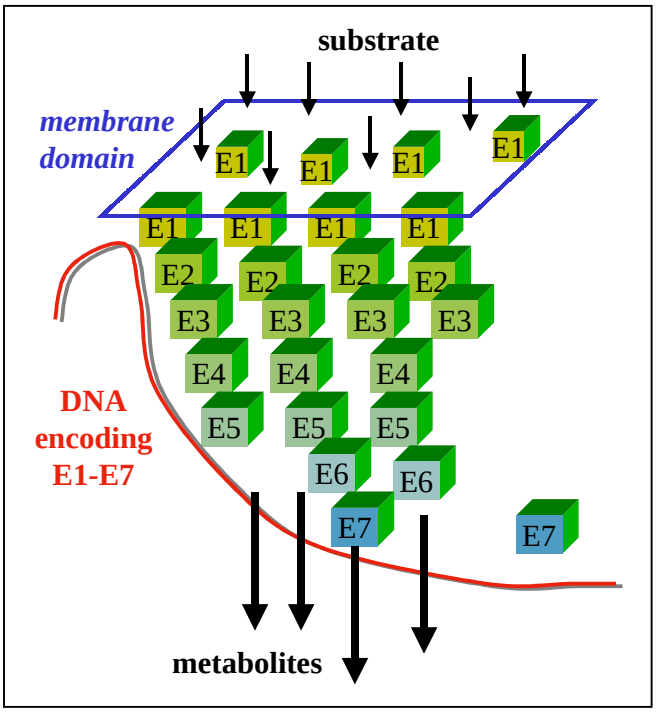


Figure 2

Horizontal links aid the assembly of a hyperstructure. Oligomeric protein E3 may bind together two identical metabolons (E1-E5 to E1-E5) or two different ones (E1-E5 to F1-F5). In the former case, E3 plays a role in the assembly of an individual hyperstructure whilst in the latter case E3 plays a role in the interaction between two different hyperstructures.

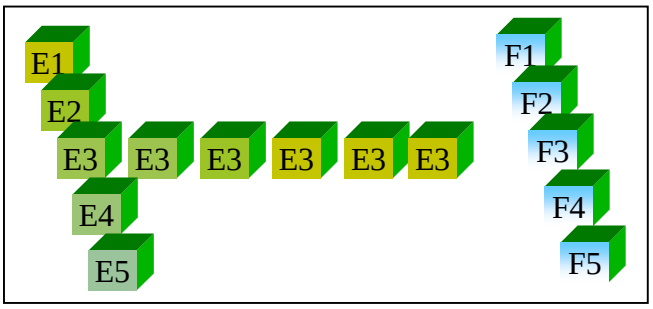


Figure 3

Local concentrations of oligomeric proteins can promote hyperstructure assembly. Protein E6 binds to its site (green) present in DNA or RNA to produce a region of the cytoplasm enriched in both E6 and its sites.

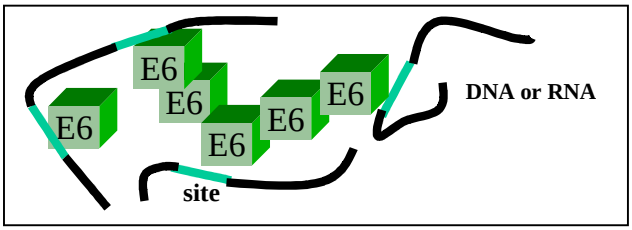


Figure 4

Transertion can nucleate hyperstructure assembly. Transertion, alias the coupled transcription, translation and insertion into and through membranes of proteins, may enrich a region of the membrane in the lipids (green) for which the proteins have an affinity. At a critical density of inserted nascent proteins, small proteolipid domains fuse into large ones and so nucleate hyperstructure assembly.

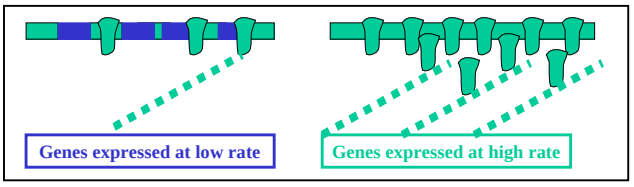


Figure 5

Cell cycle progress as a state cycle of hyperstructures. Rectangles represent non-equilibrium hyperstructures each performing one function. Blue rectangles correspond to hyperstructures with a common set of lipid (or other) preferences whilst red rectangles correspond to hyperstructures with a different set of preferences.

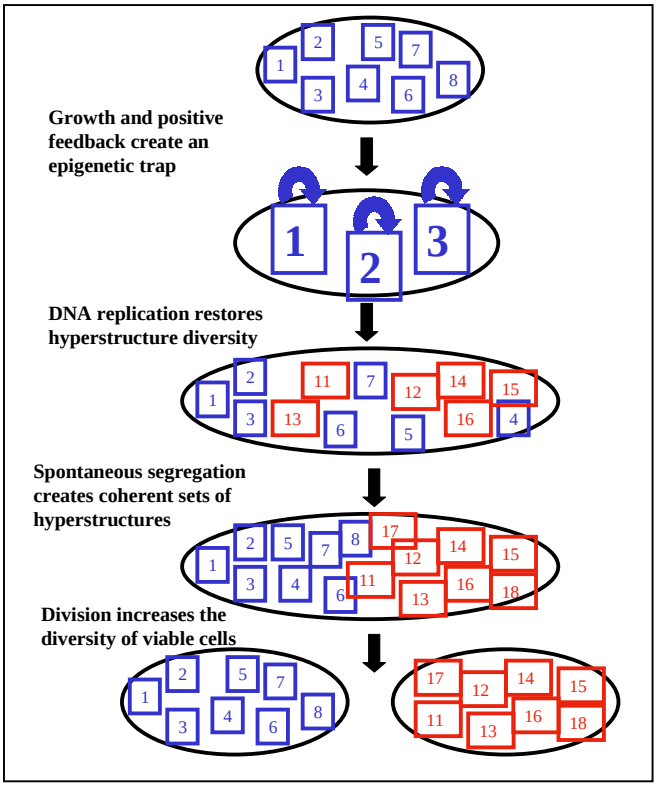


Figure 6

The spatiotemporal control of cell division by hyperstructures. The hyperstructures (the green or blue polygons) form one of two sets depending on the common preference within a set for lipids, ions, proteins etc. Each set is associated with a chromosome and is present in the future daughter cell. The division site is in the cytoplasmic membrane (thin rectangles) at the interface between these sets indicated by the arrow. Two possibilities for the structure of the division site (red arrows) which may be between the principal domains (blue and green) at either the interface or a separate, specific domain (red).

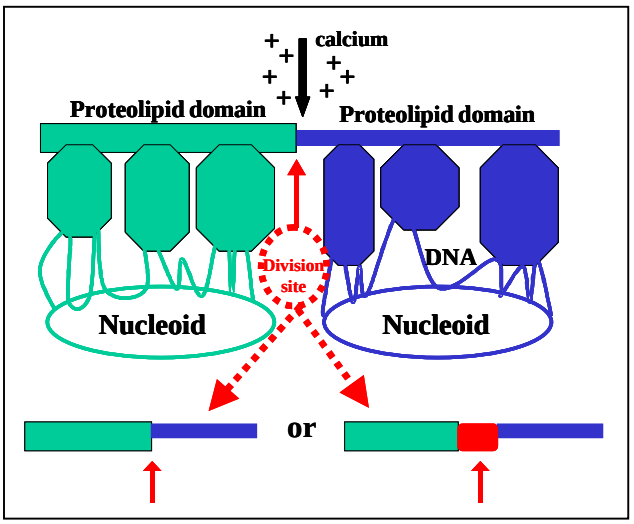


Figure 7

A two process example of parallel processing. Since each process does not need to access the space of the other one (except the boundary of the current space), each part can be stored locally in a multi-computer networked environment. The boundary of each part is the only information to be shared (i.e. transmitted between the computers).

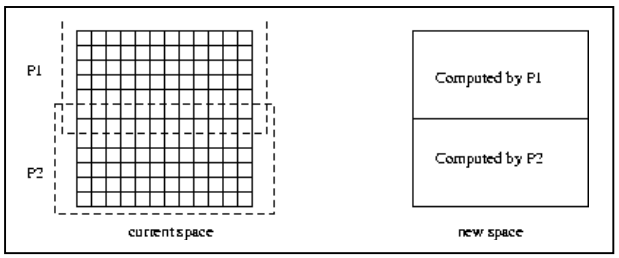


Figure 8

Movement on a 2-D grid. A molecule at time *t* can choose between 9 positions at time *t+1*. a molecule at time *t* can choose between 9 positions at time *t+1*.

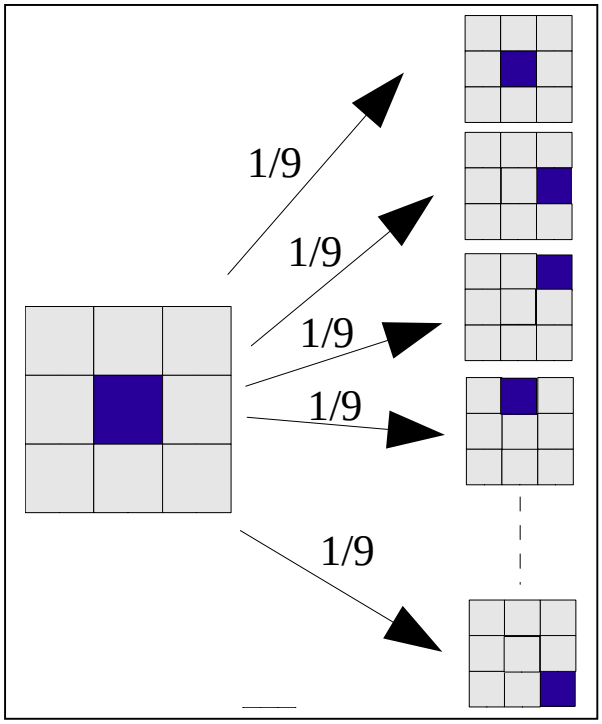
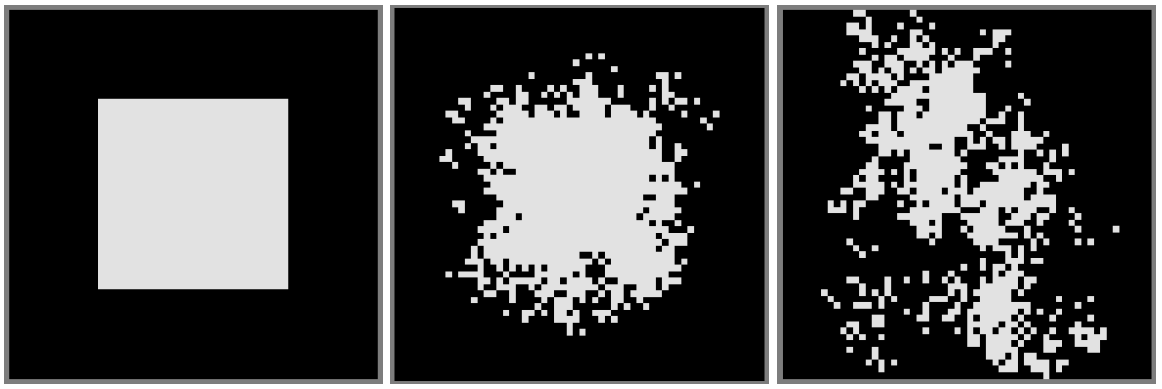


Figure 9

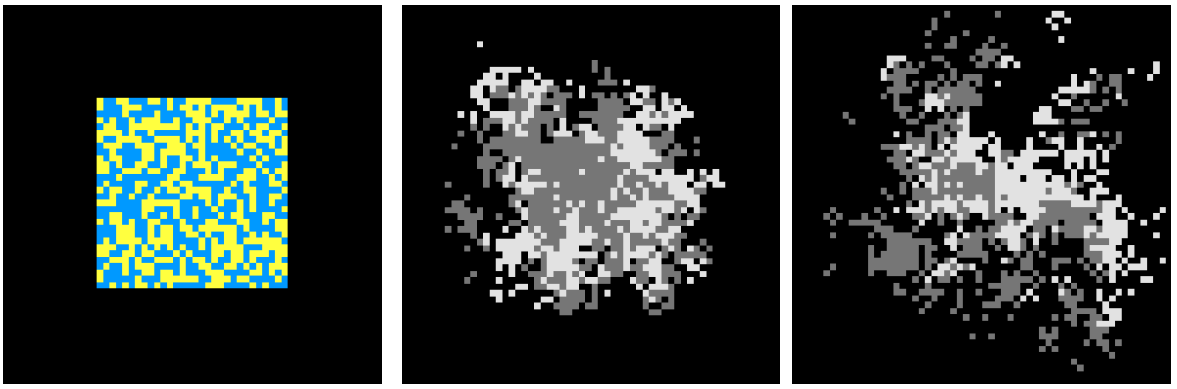
Diffusion in a cellular automata system. Empty units are black and filled units are yellow. States at successive times (t=0, 10 and 100) are shown.



t=0 t=10 t=100

Figure 10

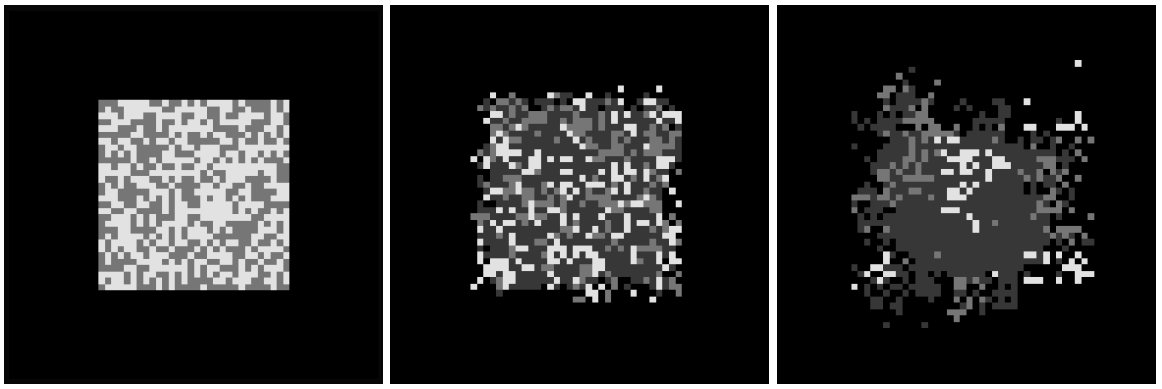
Diffusion of 2 types of molecule in a cellular automata system. States at successive times (t=0, 10 and 100) are shown



t=0 t=10 t=100

Figure 11

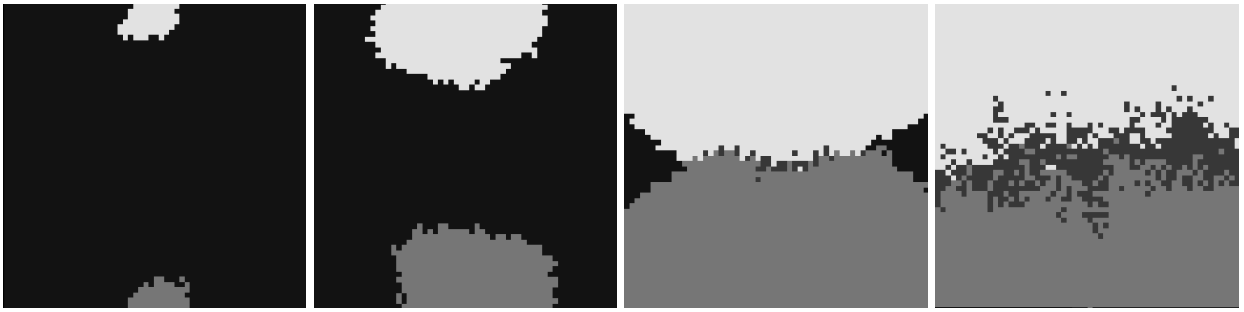
Diffusion of 2 substrates and a product in a cellular automata system. Molecules of type 1 and molecules of type 2 interact to produce type 3. At time t=0, there are only 2 types of molecules, type 1 (yellow) and type 2 (light blue). At time t=3, type 3 (dark blue) appears



t=0 t=3 t=10

Figure 12

Formation of a hyperstructure compound with 3 types of molecules. At time t=5 and t=20, 2 simple molecular structures develop. Molecules of type 1 are at the top and molecules of type 2 are at the bottom of the cellular automata. At time t=50, the two structures meet and produce molecules of type 3. Then, at time t=100, a hyperstructure with 3 types of molecules appears.



t=5 t=20 t=50 t=100