Hyperstructures, genome analysis and I-cell

Patrick Amar[1,2], Pascal Ballet[3], Georgia Barlovatz-Meimon[4], Arndt Benecke[5], Gilles Bernot[1], Yves Bouligand[6], Paul Bourguine[7], Franck Delaplace[1], Jean-Marc Delosme[1], Maurice Demarty[8], Itzhak Fishov[9], Jean Fourmentin-Guilbert[10], Joe Fralick[11], Jean-Louis Giavitto[1], Bernard Gleyse[12], Christophe Godin[13], Roberto Incitti[1], François Képès[14], Catherine Lange[15], Lois Le Sceller[8], Corinne Loutellier[15], Olivier Michel[1], Franck Molina[16], Chantal Monnier[8], René Natowicz[17], Vic Norris[8], Nicole Orange[18], Helene Pollard[19], Derek Raine[20], Camille Ripoll[8], Josette Rouviere-Yaniv[21], Milton Saier jnr. [22], Paul Soler[19], Pierre Tambourin[19], Michel Thellier[8], Philippe Tracqui[23], Dave Ussery[24], Jean-Claude Vincent[25], Jean-Pierre Vannier[26], Philippa Wiggins[27] and Abdallah Zemirline[3]*

[1]*Laboratoire de Méthodes Informatiques, CNRS UMR 8042, Université d'Evry, 91025 Evry cedex, France*

[2]*Laboratoire de Recherche en Informatique, Université Paris-Sud, Orsay, France*

[3]*Laboratoire des Interfaces Machines Intelligentes, Université de Bretagne Occidentale, Brest, France*

[4]*INSERM U492, Faculté de Médecine, 8, rue du Général Sarrail, 94010 CRETEIL, France*

[5]*Institut des Hautes Etudes Scientifiques, Bures-Sur-Yvette, France & INSERM U417, Paris, France*

[6]*Histophysique (EPHE) & Faculté de Pharmacie, 10 rue A.-Bocquel, 49100 Angers, France*

[7]*CREA, Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France*

[8]*Laboratoire des Processus Intégratifs Cellulaires, UPRESA CNRS 6037, Faculté des Sciences & Techniques, Université de Rouen, 76821, Mont-Saint-Aignan, France*

[9]*Department of Life Sciences, Ben-Gurion University of the Negev, POB 653, Be'er Sheva, Israel 84105*

[10]*Fondation Scientifique Fourmentin-Guilbert, 2 Avenue du Pavé Neuf, 93160 Noisy-Le-Grand, France*

[11]*Health Science Center, Texas Tech University, Lubbock, TX, USA*

[12]*Laboratoire de Mathématiques de l'INSA de Rouen, Mont-Saint-Aignan, France*

[13]*CIRAD, Laboratoire de modélisation des plantes, Montpellier, France*

[14]*ATelier de Genomique Cognitive, Département de Mathématiques, Université d'Évry 91025 Evry, France*

[15]*Spectrométrie de Masse Bio-organique, UFR des sciences, Université de Rouen, 76821 Mont-Saint-Aignan, France*

[16]*Faculté de Pharmacie, IBPh - CNRS UMR 5094, 15 Av. Charles Flahault, B.P. 14491, 34093 Montpellier, France*

[17]*ESIEE, Laboratoire d'Informatique, Cité Descartes, BP 99, 93162 Noisy le Grand cedex, France*

[18]*Laboratoire de Microbiologie du Froid, Université de Rouen, Evreux, France*

[19]*Genopole Research, 2 rue Gaston Cremieux, 91057 Evry, France.1057 Evry, France*

[20]*Dept. of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK.*

[21]*Institut de Biologie Physico-Chimique, CNRS UPR 9073, 75005 Paris, France*

[22]*Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA*

[23]*Laboratoire des Techniques Imagerie Modélisation Cognition, CNRS UMR 5525, Faculté de Médecine, 38706 La Tronche, France*

[24]*Center for Biological Sequence Analysis, Dept. of Biotechnology, The Technical University of Denmark, 2800 Kgs, Lyngby, Denmark*

[25]*Polymères, biopolymères et membranes, UMR 6522 CNRS, Université de Rouen, France*

[26]*Hématologie-Oncologie Pédiatrique, Faculté de Médecine, CHU de Rouen, 76813 Rouen, France*

[27]*Genesis Research and Development Corporation Limited, Parnell, PO Box 50, Auckland, New Zealand*

[*]*Author for correspondence. Email vjn@univ-rouen.fr Telephone 00 33 235 14 6908 Fax 00 33 235 14 7020*

Feldfunktion geändert

Abstract

New concepts may prove necessary to profit from the avalanche of sequence data on the genome, transcriptome, proteome and interactome and to relate this information to cell physiology. Here, we focus on the concept of large activity-based structures, or hyperstructures, in which a variety of types of molecules are brought together to perform a function. We review the evidence for the existence of hyperstructures responsible for the initiation of DNA replication, the sequestration of newly replicated origins of replication, cell division and for metabolism. The processes responsible for hyperstructure formation include changes in enzyme affinities due to metabolite-induction, lipid-protein affinities, elevated local concentrations of proteins and their binding sites on DNA and RNA, and transertion. Experimental techniques exist that can be used to study hyperstructures and we review some of the ones less familiar to biologists. Finally, we speculate on how a variety of *in silico* approaches involving cellular automata and multi-agent systems could be combined to develop new concepts in the form of an *Integrated* 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 that form membranes, and about other important molecules such as polyamines, polyphosphates and poly-ß-hydroxybutyrate. Model organisms such as the bacterium *Escherichia coli* are invaluable in making sense of this information. The 4.6 Mb genome of *E. coli* K-12 MG1655 has been sequenced (Blattner *et al*., 1997) and was found to have 4288 protein-coding genes of which 38% had no attributed function. Other strains of *E. coli* have genes that range in number from 4085 (K-12 W3110) to 5361 (O157 RIMD) whilst the bacterium *Streptomyces coelicolor* has 7846. These numbers are in the same range of those of unicellular eukaryotes such as the fission yeast, *Schizosaccharomyces pombe*, with 4824 genes and the budding yeast, *Saccharomyces cerevisiae*, with 5885. 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. Here, we define the concept of hyperstructures and give examples of possible hyperstructures in bacteria. We discuss some of the physico-chemical factors that may be involved in hyperstructure formation and we explain how the concept may help in using cellular automata and multi-agent systems to exploit the information provided by genome sequencing. Finally, we advocate an approach to the study of biological complexity *via* construction of cell *in silico* based on hyperstructures. Such a cell would be an important step towards the construction of an integrated cell, *I-cell*, that would bring together many of the processes believed to determine the structure and phenotype of real cells.

2. What are hyperstructures?

At this stage in the development of the concept, we consider a hyperstructure to be a non-equilibrium, thermodynamically open, activity-based structure. A hyperstructure is a collection of diverse molecules – genes, mRNAs, proteins, ions, lipids – that is assembled into a large, spatially distinct structure to perform a function and that is disassembled when no longer required. Certain hyperstructures may therefore structure cytoplasm, chromosome and membrane. We do not propose to try to resolve here the question of the relationship between hyperstructures and equilibrium structures.

3. Examples of possible hyperstructures

The hyperstructure concept is relevant to the organisation of both prokaryotic and eukaryotic cells but here we confine ourselves to bacteria and in particular *E. coli*. The examples below are drawn primarily from the cell cycle and from metabolism. It should become clear that the archetypal bacterial hyperstructure contains genes being

transcribed, their mRNAs being translated and the nascent proteins being inserted into the membrane to form a domain with a characteristic lipid composition (Figure 1).

3.1 Cell cycle.

The cell cycle comprises the initiation of DNA replication and the temporary sequestration of the newly replicated origins of replication (to prevent multiple rounds of replication), the separation and segregation of the chromosomes into the future daughter cells, and the division of the parental cell between the chromosomes, probably followed by the inactivation of the division apparatus.

An initiation/origin of replication hyperstructure. This comprises the DnaA protein and certain of the sites on DNA to which it binds. In *E. coli*, the ATP-bound form of DnaA is required for initiation of replication *in vitro* whilst the ADP-bound form is inactive (Castuma *et al*., 1993). Both ATP-DnaA and ADP-DnaA bind to 9mer DnaA boxes, TTA/TTNCACA, but only ATP-DnaA protein binds in addition to a 6mer site, AGATCT (Speck *et al*., 1999). DnaA is associated with the membrane *in vivo* (Newman and Crooke, 2000) and initiation requires a membrane with a particular phospholipid composition (Fralick and Lark, 1973) and domain structure to activate DnaA (Castuma *et al*., 1993; Xia and Dowhan, 1995) as well as the polymerisation of DnaA (Weigel *et al*., 1999). In one hypothesis, this putative hyperstructure has a transient existence that depends on the dynamics of other hyperstructures. It is clearly a non-equilibrium structure, indeed, it is disassembled after accomplishing the act of initiation (Norris *et al*., 2002).

A DNA replication hyperstructure. This comprises the SeqA protein, key enzymes in DNA replication, and the genes that encode them (Norris *et al*., 2000). SeqA sequesters newly replicated origins and is found in clusters (Onogi *et al*., 1999). It binds to GATC sequences and it polymerises. The non-equilibrium nature of this hyperstructure results from its dependence on the energy-consuming process of replication with SeqA binding preferentially to hemi-methylated GATC sites in genes encoding enzymes responsible for DNA replication, topology, repair and precursor synthesis such as ribonucleoside diphosphate reductase (Guzman *et al*., 2002).

A cell division hyperstructure. This comprises the ten or so division proteins plus enzymes involved in peptidoglycan synthesis plus the genes that encode them (Buddelmeijer *et al*., 1998; Norris and Fishov, 2001). Many of these genes are located together and are transcribed together in the *dcw* cluster at the 2 min position on the chromosome. This is a non-equilibrium structure in that it requires the energy-consuming processes of transcription and translation to bring these genes together at the membrane along with the nascent proteins (see below).

3.2 Metabolism

A secretion hyperstructure. 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 (ABC protein) which then binds to PrtE (membrane fusion protein) and which binds to PrtF (outer membrane protein) (Letoffe *et al*., 1996).

A glycolytic hyperstructure. The glycolytic pathway can be extracted as an equimolar complex of 1.65 megaDa that reveals compartmentation of substrates (Mowbray and Moses, 1976). Evidence has also been obtained for the existence of metabolons which are assemblies of the enzymes that act in succession in a pathway (Mitchell, 1996; Velot *et al*., 1997). Of course, such metabolons, which may or may not be non-equilibrium structures (see below), may themselves associate into larger structures or hyperstructures. The jury is still out on the question of whether small assemblies of enzymes should be considered as hyperstructures or whether this term should be reserved for a single, large hyperstructure that would group together genes, mRNA and enzymes and that would structure chromosome, cytoplasm and membrane.

A PTS hyperstructure. 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 P from phosphoenolpyruvate; E2s+E1+HPr are proposed to form a complex in response to the presence of the appropriate substrate as part of a 'metabolite-induced metabolon' event (Norris *et al*., 1999). The idea is that a PTS hyperstructure forms when the cell is actively engaged in processing substrates and disappears when it is not. (We shall discuss elsewhere the exact relationship between metabolons and the full-blown hyperstructure.) Formation of a PTS hyperstructure 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).

3.3 Other hyperstructures.

These include a DNA compaction hyperstructure possibly involving the MukB protein which can form foci (Ohsumi *et al*., 2001); a nucleolus-like hyperstructure for ribosome synthesis (Lewis *et al*., 2000; Woldringh and 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 and Levit, 2000) plus, we propose, the genes encoding the abundant chemotaxis proteins (and hence the necessity for hydrolysis of ATP and GTP during transcription and translation).

4. Examples of physico-chemical and other factors that may be involved in hyperstructure formation

There are several complementary possibilities:

4.1 Protein-protein affinities

The idea here is that successive enzymes in the same pathway can be activated by their substrates to bind to one another in a heteropolymeric organisation that is sequential (or *vertical* insofar metabolic pathways are often drawn vertically with metabolites entering at the pathway at the top of the page and products as leaving at the bottom). A complementary idea is 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 homopolymeric or *horizontal* organisation could help nucleate and stabilise hyperstructures (Figure 2).

4.2 Protein-lipid affinities

It can be argued that proteins with pronounced affinities for specific lipids may congregate with those lipids in a positive feedback fashion to form the membrane domain part of a hyperstructure. There is evidence that concomitant with overproduction of a membrane protein there is a compensatory overproduction of the lipid for which it has an affinity (Arechaga *et al*., 2000). (Note that this suggests a semi-automated, general strategy might be developed in which different peptides are overproduced to derive consensus sequences to be used to interpret the genome and construct a 'lipidome'.)

4.3 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) and the binding of DnaA to its sites. 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) (Figure 3).

4.4 Steric effects

It has been suggested that enzymes in complexes are more likely to escape proteolytic degradation than when 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.

4.5 Water preferences

Water exists as nm-sized microdomains of differing structure and density that must have different solvent properties (Robinson *et al*., 1999; Wiggins, 1990). The difference in density appears to be of the order of 30% (Cho *et al*., 1997). Such differences should affect the distribution and activity of cellular constituents near, for example, the surfaces of membranes (Mayer and Hoppert, 1996). An important but difficult question is the extent to which the water preferences of the constituents of hyperstructures might determine hyperstructure formation and interaction.

4.6 Transertion

Transertion is the coupled transcription, translation and insertion into and through membranes of proteins (Binenbaum *et al*., 1999; Lynch and Wang, 1993). 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 and Brusilow, 1993), would result in assembly of an ATP synthesis hyperstructure.

5. A coherent phenotype *via* collaboration between hyperstructures

An individual hyperstructure is a structure that forms as a result of an activity. Reciprocally, a collection of hyperstructures confers the phenotype on the cell in which case the activity of the cell stems from its constituent (hyper)structures. In our hypothesis, a collection of hyperstructures also confers a coherent phenotype in which, for example, enzymes appropriate for growth in cold oxygenated conditions are 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.

5.1 Coherence via competition between hyperstructures

In an activity-based vision of the cell, only a subset of its constituents – such as those found in a hyperstructure – 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). To illustrate the concept, consider the problem of selecting an amateur football team each week from a larger group of potential players. One important factor in deciding who plays *next* week is who plays this week. This is a *status quo* process at work. It is, for example, easier to discuss shared transport arrangements with those already present than with those who are not playing that week. Hence those who play this week are mostly likely to play again next week. Another important factor is the coherence of the team. Suppose there are two candidate goalkeepers available next week (and a team is only allowed one goalkeeper), the choice of only one of them may have consequences on the choice of other players since players must be chosen who can play together (and perhaps travel together). This is a *coherence* process at work. In this analogy, the composition of the team each week is determined by a competition between two sorts of processes so as to satisfy the demands of both the *status quo* and coherence.

5.2 Coherence might be achieved by hyperstructures sharing binding proteins

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 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 and 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 and 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 Using cellular automata to test the hyperstructure hypothesis

In silico experiments prove to be a good preliminary step to test the value of the hyperstructure approach in interpreting genomic and biochemical data.

6.1 Cellular automata

Brute force computations starting from known properties of atomic interactions within macromolecules or between macromolecules would lead to an intractable problem: a powerful computer, solving differential equations, would require about one week of computation to simulate one nano-second of the behaviour of a protein. As mentioned in section 4, biochemistry deals with the higher level of the interactions between macromolecules that, moreover, take effect only when the two molecules are close enough. This allows more abstract computations to be performed with a greater efficiency. Hence, cellular automata can be particularly suitable for modelling the dynamics of interactions between molecules in 3 dimensions (Vichniac, 1984).

In a cellular automaton, the 3 dimensional space is partitioned into small unit volumes, called *voxels*, each of which may contain a macromolecule of any type. Once the automaton have been assigned an initial state (the content of each voxel), the evolution of this states is updated at a given time step. For each voxel, the update criterion depends on the state of the 26 neighbouring voxels.

6.2 Metabolite induction

To estimate the values of the parameters governing the formation of hyperstructures in bacteria, we have constructed a preliminary version of a cellular automaton that simulates the dynamics of the localisation of the phosphotransferase system (PTS) – responsible for the uptake and sensing of many sugars – and glycolytic enzymes in both a 2 dimensional membrane and a 3 dimensional cytoplasm (Le Sceller *et al*., 2000). Within a volume of $8\mu m^3$, partitioned into 200×200×200 voxels, which is more than sufficient to represent *E. coli*, we simulated a bacterium containing 5 types of cytoplasmic enzymes (3000 copies of each) and 1 type of membrane receptor (2000 copies). When the affinities of enzymes for one another were realistically increased in the presence of substrates, there was a reorganisation of both membrane and adjacent cytoplasm into structures, each of which contained up to 500 enzymes.

6.3 Data structures, description language and algorithms

A new kind of cellular automaton is under development whose main purpose is to take benefit of abstraction and local rules to improve efficiency. It automatically detects hyperstructures and allows 3 dimensional browsing while the simulation is running. Recent results from research in computer science have been used to:

- maintain low cost representation of the space with a sophisticated data structure;
- design an independent, legible, domain-oriented language to describe the pairwise interactions of molecules; it allows for example modification of the interaction rules in a way clearly separated from the internal machinery of the automaton;
- design a fast algorithm, mainly based on a uniform random choice of the pairs of molecules susceptible to interact, considerably reducing the simulation time; moreover this algorithm can be easily distributed on a computer network;

The combination of these computer science techniques allowed us to operate with very thin time slices. Roughly speaking, this strategy consists in applying elementary operations at a high rate. We call a *generation* the processing of all the molecules present in the considered space. Hence successive generations are computed very quickly.

A typical simulation scenario begins with the description of the molecular interactions. For example, the elementary interactions between an enzyme and its substrate can be described as follows (comments are introduced by //)

```
E + S <-> E * S [0.5, 0.01];  // enzyme E and substrate S make the complex ES with
                              // probability 50%, and dissociate with probability 1%
E * S <-> E * P [0.6, 0.02];  // enzyme E catalyses the transformation of substrate S to
                              // product P with probability 60%, and the reverse reaction with
                              // probability 2%
E * P <-> E + P [0.8, 0.01];  // enzyme E releases the product P
```

The next step is to define the initial state of the simulation. Easy-to-use primitives are provided, for example: **`cube (0, 0, 0, 8, E)`** centres a 8×8×8 cube (containing 512 copies of enzyme E) at the origin (0, 0, 0).

The last step is to start the simulation, during which it is continually possible to visualise any part of the 3 dimensional space (at any scale and from any point of view) with distinctive colours for each molecule (Figure 5). The system is able to compute for example 12000 generations per minute with about 4000 molecules, independently of the number of rules, and of the size of the space. If we add well-chosen rules describing affinities between

enzymes in presence of their substrates, we can observe the genesis of the first hyperstructures after about 2000 generations.

6.4 Manipulating hyperstructures

The *****-links between molecules are dynamically maintained and updated in the data structure by the automaton. Classical algorithms to extract connex components of a graph are used to recognise hyperstructures. The automaton techniques described so far are appropriate for analysing the process of assembly and disassembly of hyperstructures as described in section 4. When the hyperstructure reaches a certain size, the local processing between molecules performed by the automaton becomes unsuitable for computing the global behaviour of the hyperstructure. At this level, multi-agent systems become a better computational paradigm to implement the interactions between hyperstructures as described in section 5.

6.5 Using experiments to validate the model

The principal, immediate task is to design and conduct experiments to determine all the parameters – and the values of these parameters – that are involved in the computations described above. Obvious parameters are the choice of the interaction rules and their probabilities. In standard cellular automata, two other parameters that can have a surprisingly important impact on the results of the simulations are the choice of the time step and the spatial granularity. The remarkable efficiency of the automaton algorithm permits many simulations with numerous values for both the time step and the spatial granularity and shows that these parameters have no significant impact on the observed results. Thus the key parameters are the interaction rules. Many of these rules can be deduced and quantified from existing biochemical data. A good approach to determining the values of the probabilities to be used in simulating the interactions between macromolecules is to simulate well-characterised *in vitro* experiments and to adjust the values until the simulation gives the experimentally observed results, that is, to use the simulation itself to determine the values rather than to interpret and import values directly from the literature. Finally, of course, the results of the simulation should be tested against experimental reality and this requires experimental techniques to observe real hyperstructures *in vitro* or *in vivo* (Section 8).

7. Equilibrium structures

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 in the *E. coli* K-12 MG1655 genome 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 which binds preferentially to unusual structures such as kinked or cruciform DNA (Bonnefoy *et al*., 1994; Kamashev and Rouviere-Yaniv, 2000). Equilibrium structures include immiscible domains within condensed chromosomes where much of the DNA is probably in a cholesteric form (Livolant and Leforestier, 1996). This immiscibility occurs in the context of an ordered liquid, with the DNA closely layered by a regular twist (Bouligand and Norris, 2001), a situation that may minimize entangling and facilitate co-expression of the genes within a domain. The challenge is to translate this information into a dynamic 3-D model. One model that might be tested *via* cellular automata (see above) 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.

8. Experimental aspects

In parallel with modelling hyperstructures, it is essential to develop or apply experimental techniques to investigate them. These fall into two classes, those that can reveal hyperstructures formed *in vivo* and those that can reveal the factors underlying hyperstructure formation and operation *in vitro*.

8.1 Detecting hyperstructures formed in vivo

Visualising hyperstructures directly with conventional techniques has been difficult since it requires the co-localisation 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 *et al*., 1993). 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.

8.2 Revealing the effects of hyperstructure formation and operation in vitro

Studying hyperstructure formation is hampered, in the case of glycolysis, by a shortage of 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 that might be used (Gygi *et al*., 1999). More seriously, we lack details of the constants of affinity of the PTS and glycolytic enzymes. These could be obtained using optical waveguide lightmode spectroscopy 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).

Studying how hyperstructures might create membrane domains to attract and activate FtsZ and hence initiate cell division requires, ideally, the construction of an *in vitro* division system. 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 just such a division system (Alexandre *et al*., 2002). 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 also be used to try to construct an *in silico* model of the formation of a division hyperstructure.

9. Speculations: from H-cell to I-cell

Developing new concepts may prove essential to a full understanding of how a cell works. To test and develop such concepts, we have advocated here the construction of a cell *in silico* based on hyperstructures that might be termed an *H-cell*. It would, however, only treat part of the myriad processes and molecules that must be modelled if we are ever to fully grasp cellular reality. H-cell is but a step toward an *Integrated* cell – *I-cell* – that would have the ambition of modelling much of what is found or is suspected in real prokaryotic and eukaryotic cells. The I-cell we envisage would undergo selection for growth, differentiation, communication with other cells and survival in a world of artificial chemistry in which hyperstructures could play a key role (Dittrich and Banzhaf, 1998). We are currently exploring a version of this in which artificial bacteria comprising autocatalytic networks of 'enzymes' *alias* sequences of numbers grow on 'amino acids' *alias* individual numbers (Demarty *et al.*, submitted). In a more sophisticated version, 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 and 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. In addition, an I-cell might offer a way to discover the importance of a particular organising process, for example, one based on water structure or tensegrity or reptation (the constrained movement of polymers in a crowded solution). 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 Jean-Marc Valleton for comments and Genopole and the Conseil Régional de l'Ile de France for support. AB is supported by a grant from the Deutsche Forschungsgemeinschaft. DU was funded by a grant from the Danish National Research Foundation.

Reference

Alexandre, S., G. Cole, S. Coutard, C. Monnier, V. Norris, W. Margolin, X. Yu and J.-M. Valleton (2002). Interaction of FtsZ protein with a DPPE film. Colloids and Interfaces B 23: 391-395.

Arechaga, I., B. Miroux, S. Karrasch, R. Huijbegts, B. de Kruijff, M.J. Runswick and J.E. Walker (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., L. Claret, R. Hengge-Aronis and J. Rouviere-Yaniv (2001). The *Escherichia coli* histone-like protein HU regulates *rpoS* translation. Molecular Microbiology 39: 1069-1079.

Binenbaum, Z., A.H. Parola, A. Zaritsky and I. Fishov (1999). Transcription- and translation-dependent changes in membrane dynamics in bacteria: testing the transertion model for domain formation. Molecular Microbiology 32: 1173-1182.

Blattner, F.R., G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau and Y. Shao (1997). The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453-1462.

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

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

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

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

Buddelmeijer, N., M.E.G. Aarsman, A.H.J. Kolk, M. Vicente and N. Nanninga (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., E. Crooke and A. Kornberg (1993). Fluid membranes with acidic domains activate DnaA, the initiator protein of replication in *Escherichia coli*. Journal of Biological Chemistry 268: 24665-24668.

Cho, H.C., S. Singh and G.W. Robinson (1997). Understanding all of water's anomalies with a non-local potential. Journal of Chemical Physics 107: 7979-7988.

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

Fralick, J.A. and K.G. Lark (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.

Guzman, E.C., J. Caballero, L. and A. Jimenez-Sanchez (2002). Ribonucleoside diphosphate reductase is a component of the replication hyperstructure in *Escherichia coli*. Molecular Microbiology 43: 487-495.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb and R. Aebersold (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 17: 994-999.

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

Ksenzenko, S.M. and W.S.A. Brusilow (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., C. Ripoll, M. Demarty, A. Cabin-Flaman, T. Nyström, M. Saier Jnr. and V. Norris (2000). Modelling bacterial hyperstructures with cellular automata. Interjournal Paper 366: http://www.interjournal.org.

Letoffe, S., P. Delepelaire and C. Wandersman (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., S.D. Thaker and J. Errington (2000). Compartmentalization of transcription and translation in *Bacillus subtilis*. EMBO Journal 19: 710-718.

Livolant, F. and A. Leforestier (1996). Condensed phases of DNA: structures and phase transitions. Progess in Polymer Science 21: 1115-1164.

Lynch, A.S. and J.C. Wang (1993). Anchoring of DNA to the bacterial cytoplasmic membrane through co-transcriptional synthesis of polypeptides encoding membrane proteins or proteins for export : a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I. Journal of Bacteriology 175: 1645-1655.

Mayer, F. and M. Hoppert (1996). Functional compartmentalization in bacteria and archaea. A hypothetical interface between cytoplasmic membrane and cytoplasm. Naturwissenschaften 83: 36-39.

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 V. Moses (1976). The tentative identification in *Escherichia coli* of a multi-enzyme complex with glycolytic activity. European Journal of Biochemistry 66: 25-36.
Newman, G. and E. Crooke (2000). DnaA, the initiator of *Escherichia coli* chromosomal replication, is located at the cell membrane. Journal of Bacteriology 182: 2604-2610.
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., M. Demarty, D. Raine, A. Cabin-Flaman and L. Le Sceller (2002). Hypothesis: hyperstructures regulate initiation in *Escherichia coli* and other bacteria. Biochimie 84: 341-347.
Norris, V. and I. Fishov (2001). Hypothesis: membrane domains and hyperstructures control bacterial division. Biochimie 83: 91-98.
Norris, V., J. Fralick and A. Danchin (2000). A SeqA hyperstructure and its interactions direct the replication and sequestration of DNA. Molecular Microbiology 37: 696-702.
Norris, V., P. Gascuel, J. Guespin-Michel, C. Ripoll and M.H. Saier Jr. (1999). Metabolite-induced metabolons: the activation of transporter-enzyme complexes by substrate binding. Molecular Microbiology 31: 1592-1595.
Norris, V. and L. Le Sceller. (2001). *International Conference of Systemics, Cybernetics and Informatics, Orlando, Florida, USA*.
Ohsumi, K., M. Yamazoe and S. Hiraga (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., H. Niki, M. Yamazoe and S. Hiraga (1999). The assembly and migration of SeqA-Gfp fusion in living cells of *Escherichia coli.* Molecular Microbiology 31: 1775-1782.
Ouzounis, C., G. Casari, A. Valencia and C. Sander (1996). Novelties from the complete genome of *Mycoplasma genitalium*. Molecular Microbiology 20: 898-900.
Pedersen, A.G., L.J. Jensen, S. Brunak, H.-H. Staerfeldt and D.W. Ussery (2000). A DNA structural atlas for *Escherichia coli.* Journal of Molecular Biology 299: 907-930.
Ramsden, J.J. (1993). Review of new experimental methods for investigating random sequential adsorption. J. Statist. Phys. 73: 853-877.
Revet, B., B. von Wilcken-Bergmann, H. Bessert, A. Barker and B. Müller-Hill (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., B. Gontero, L. Avilan and S. Lebreton (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., C.H. Cho and J. Urquidi (1999). Isobiestic points in liquid water: Further strong evidence for the two-state mixture model. Journal of Chemical Physics 111: 698-702.
Speck, C., C. Weigel and W. Messer (1999). ATP- and ADP-DnaA protein, a molecular switch in gene regulation. EMBO Journal 18: 6169-6176.
Stock, J. and M. Levit (2000). Signal transduction: hair brains in bacterial chemotaxis. Current Biology 10.: R11-4.
Thellier, M., C. Ripoll, C. Quintana, F. Sommer, P. Chevallier and J. Dainty (1993). Physical methods to locate metal atoms in biological systems. Methods in Enzymology 227: 535-586.
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., T.S. Larsen, K.T. Wilkes, C. Friis, P. Worning, A. Krogh and S. Brunak (2001). Genome organisation and chromatin structure in *Escherichia coli.* Biochimie 83: 201-212.
Velot, C., M.B. Mixon, M. Teige and P.A. Srere (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., A. Schmidt, H. Seitz, D. Tungler, M. Welzeck and W. Messer (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 N. Nanninga. (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,
Xia, W. and W. Dowhan (1995). *In vivo* evidence of the involvement of anionic phospholipids in initiation of DNA replication in *Escherichia coli*. Proceedings of the National Academy of Science U.S.A. 92: 783-787.

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.

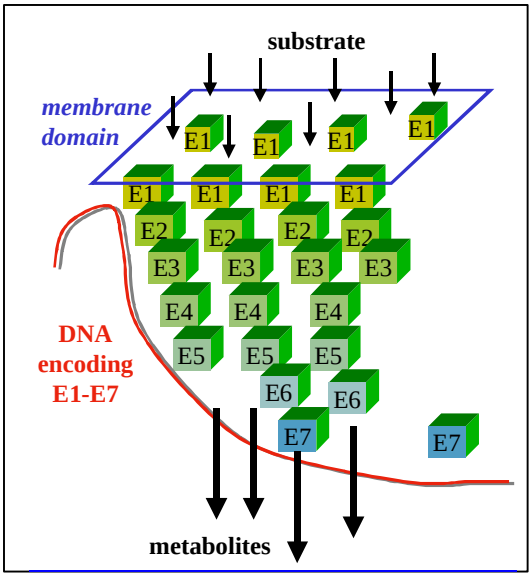

Figure 2

Horizontal or homopolymeric 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.

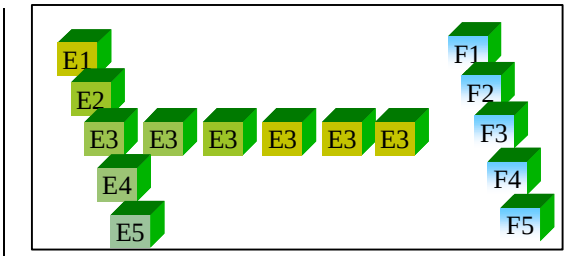

Figure 3

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

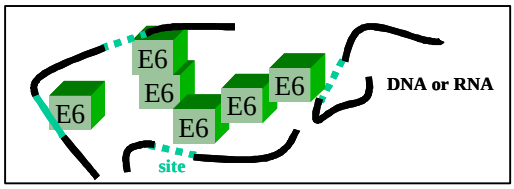

Figure 4

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. The dotted lines represent mRNA, the asymmetrical objects represent proteins and the contiguous dark and light grey rectangles at the top of the figure represent different phospholipids.

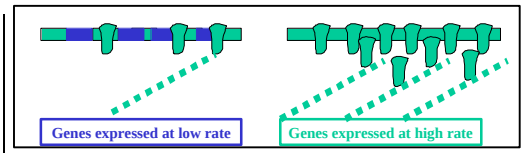

Figure 5

Formation of a hyperstructure due to changes in the affinity of its constituent enzymes for one another. In this simulation, the cell membrane is represented by a sphere, the membrane receptors are displayed in light blue (protA) and may be linked to two copies of the enzyme A (enzAL) whilst all the other enzymes may only be linked to one other enzyme (this example is very similar to the one shown in Figure 1).