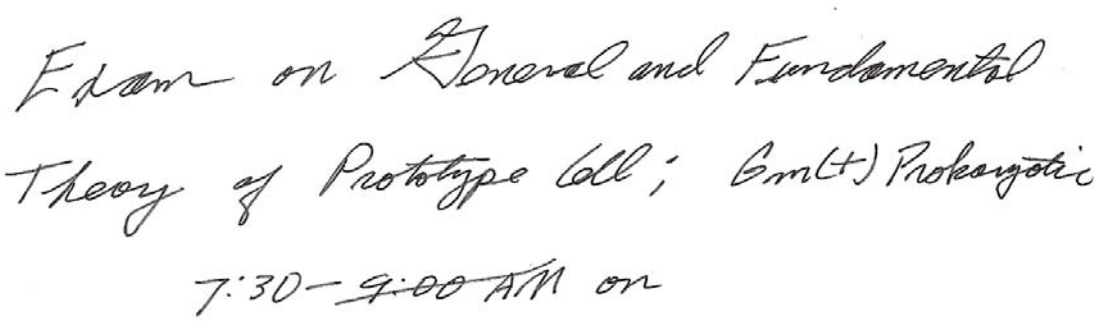
What I was doing when I was Einstein's age of discovery

David Martin Degner
Degner Scientific and Engineering
Anchorage, Alaska
davidmartindegner@gmail.com

Abstract: When I was in my twenties I worked on the problem: What is Life? In 1980/81 I was a research associate in molecular genetics at the University of California at Berkeley working for a professor in bacteriology. I wrote for the professor a 57 legal sized page, hand written and unedited essay in the form of a sequence of 72 questions and answers that focused on Gram (+) prokaryotic cells, the hydrogen of biology. The professor failed to comment or ask a single question. Back then I was into equilibrium and non-equilibrium thermodynamics, reversible and irreversible processes, chemical interactions and kinetics, static and dynamic phase organization, mass/energy and information flow, information definition and information processing, and modeling biological cells as chemical computers, i.e. Turing machines. I've faithfully transcribed that unedited sequence of questions and answers here. It's at times quite tedious and there were some major errors but there also is some really good science. There is always chaos embedded in new work. This is a qualitative, phenomenological model of the central processing unit of the prokaryotic cell that is the first step in obtaining a fully quantitative model for biology.

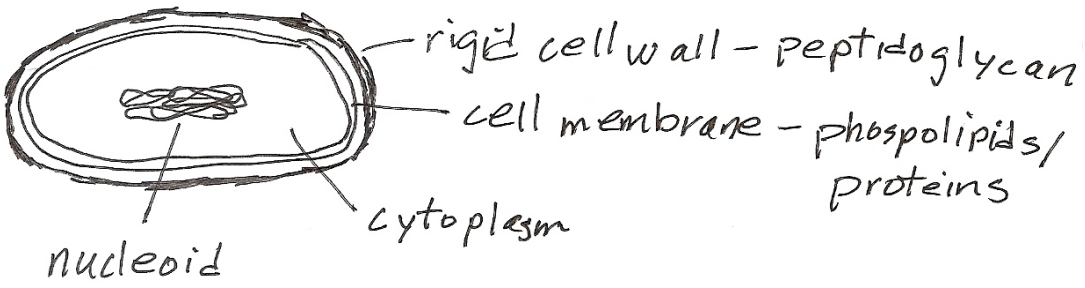


1. What is the Gram (+) cell?

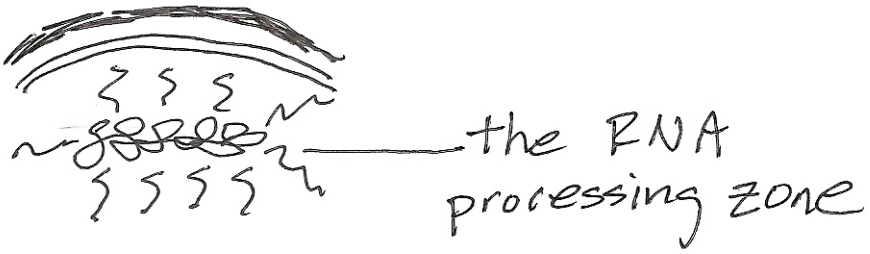
   The cell is a collection of molecules and atoms of several classes

   a) nucleic acid polymers
   b) amino acid polymers
   c) sugar or carbohydrate polymers
   d) phospholipids
   e) small amtsounts of copolymers of the above
   f) small molecule precursors of the above
   g) H_2O, ions, H^+

2. What is the state or stationary phase organization of the gm (+) cell?



3. What additional phase obtains in the growing cell?



4. What is meant by a phase and how does this affect transport?

   A phase is any region of space over which the composition is constant net of thermal fluctuations. This is a definition of and continuity requirement for any phase. Within a phase in the absence of a flow system transport is strictly limited to random walk diffusion.

5. What is a dissipative structure?

   There are two kinds: Those motivated by energy exchange across a boundary (Benard instability) and those motivated by a chemical reaction. The latter type is of interest here. These are spatial temporal organizations of matter such that the chemical relaxation of the species following the electronic transitions motivate, this transport replenishing reactants and removing products from the reaction which drives the system. Their continued operation in time requires the continuous addition of reactants and removal of products. An ideal dissipative structure would solve this transport problem by converting products → reactants, an uphill reaction since this is the source of chemical bond energy which provides for the irreversibility of the flow system.

6. Define the nucleic acid mainframe.

There are three components – DNA ds (double strand), RNA ss (single strand) and the d & r NTP's.

It is a dissipative structure.

The DNA is condensed into a symmetric volume which the monomers can diffuse into. Synthesis of RNA from monomers occurs throughout the condensed DNA. The RNA ss migrates out. This is a vectorial flow process. It is driven by the relaxation of the RNA ss from a constrained configuration within the DNA matrix where 2°/3° (secondary/tertiary) structure cannot obtain to the unconstrained structure obtainable at the surface of the nucleoid where 2°/3° forms. Polymers fold very fast, and if the ss was not constrained it would fold right away.

7. How does the RNA get to the surface of the nucleoid from its point of synthesis within the nucleoid?

   It does a random walk assisted by gross mixing of the DNA ds, but is primarily a random walk. There is no other way because of the phase definition. The DNA ds undergoes gross mixing in the nucleoid such that all DNA in the nucleoid is equivalent, save for the localization by immobilization by high transcriptional activity implying a highly constrained DNA matrix in that region because of interwound ss. This mixing allowing equivalence of all DNA is important so some DNA does not get left on the surface where it may not be transcriptionally active due to the high thermal noise at the interface and different conc. environment.

8. How are the nucleic acid monomers partitioned between cytoplasm/nucleoid?

   This is a very interesting question, the answer not being entirely clear. It is very interesting on how the partitioning is affected by the phosphorylation level, and metal ion complexation, and degree of protonation.

   To separate the d & r NTP's from the cytoplasm to have available in higher conc for the polymerization reaction requires work.

   This separation work is spread evenly over the nucleoid surface, allowing the maximal interfacial surface area available where the separation work will be done.

   In classical theory, the partitioning of molecules between phases is said to occur such that the chemical potential of all species is identical throughout the system. This is of low utility because nothing is provided by way of mechanism or kinetics for the partitioning process and the term "chemical potential" is simply a catch all for all we don't know but would like to. It balances the books and makes things nice and neat but provides no useful knowledge or information..

To understand the partitioning we will have to ask about the interactions of the species in the two phases, also the diffusivity or mobility.

The d & r NTP's must have high mobility through the nucleoid to allow their transport to the reaction centers. Since the DNA ds is a continuous sequence of liganding sites they can interact frequently and at many locations leading to higher conc. there.

An interesting question concerns the actual selection at the interface with the cytoplasm. Differential diffusivity and steric effects play an important role here.

Breaking the phosphodiester bonds in the nucleoid assoc with nucleic acid synthesis leads to using up H_2O and releasing H^+. The high mobility of H^+ means they will escape fast from the site of synthesis. Mg^{++} will probably partition to fill the (–) electric void left behind. Enhanced complexation of NTP's with Mg^{++} at the interface may work in some way as a filter for admission. This is speculative.

9. What is the Thermal Code?

The DNA has a general, static 2° structure. The dynamic aspects are not generally well understood. 2° conformation implies 3° also by connectivity definition. So the 2°/3° dynamic conformation is poorly understood.

The Thermal Code is based on the concept that the primary recognition event for interaction between DNA molecule and proteins is motivated by some 2°/3° signal in the DNA molecule. Certain 2°/3° are the limiting substrates for the interaction.

The Thermal Code construct asserts that the 2°/3° signal channel for interaction with proteins is a function of the DNA sequence and the monomer concentration.

The interaction of monomers with the ds by transient H-bonding in the major groove is what couples the environment to memory, is the primary interface for exchange of information in the system, it is the deepest level of communication in the system, it the heart of the phenotype/genotype information transfers.

It requires that this interaction leads to some conformational property which determines the DNA/protein interaction profile.

The DNA sequence by itself exhibits reasonably well known thermotropic behavior. The monomer interaction brings about lyotropic behavior.

There must be meaningful coupling between the rate of formation of signals as a *f*(sequence, monomers) and the utilization of signals through transcription initiation, propagation, termination and replication initiation.

Transcription initiation is a slow event in the microscopic world – no faster than one every 1 – 2 sec. For molecular signals to be generated on this time scale is a nontrivial problem. It requires long-range cooperative with a very large polymer in a very specific environment. It requires that the polymerization reaction itself be a relatively slow chemical event, which transcription is at 50 – 75 bases/s. This allows relaxation of the DNA 2°/3° following (behind) polymerization to be slow.

10. What is the abstract information capacity of such an array/monomer concentration system?

    Essentially ∞. The theoretical limit of information capacity in the monomer concentration is established by three factors

    1) the solubility limit of the monomers (competition between monomers is allowed)
    2) the number of different monomers possible
    3) the ability to convert different concentration profiles into a recognizable event

    This last limitation will be a strict function of the # of liganding or interaction sites available and the mechanism of summing over these interaction sites to produce a recognizable 2°/3° conformational signal. The theoretical limit of the array will be a function of the array length, for DNA going up as 4^N, very large! Persistence length will be the absolute limit of possible communication along the array molecule.

11. What is the actual nature of the signal and the molecular mechanism for its generation?

    The only thing which can be asserted unequivocally is that upon collision of RNA polymerase with the DNA molecule some complex is formed if the signal is present. This would be the primary initiation event for transcription. Additional sequence specific interactions may be required as secondary events, such as interaction with upstream or Pribnow homology. Tertiary events may be binding of the first several bases to this 2° initiation complex.

    It is asserted in the Thermal Code model that the central regulatory event is the primary initiation event of getting into the helix, and the 2° event of specific sequence recognition serve to define a clean start for polymerization and strand selection. The 3° event of binding the first several bases may be a volume regulation, if 2 bases are important a complex satisfying the first two sequential initiation events could have 16 possible rates.

12. What is the role of 2°/3° structure of the ss at the nucleoid/cytoplasm interface?

    The 2°/3° for RNA determines the interaction with ribosomes, regulatory proteins, nucleases and complexation with ions and small molecules. This is the maturation

process for stable RNA and translational control for messenger RNA. Since the RNA is coming out of the nucleoid as it is synthesized, the folding is defined sequentially. Competition for different structures and for specific sequences effect orderly kinetic processing and regulation.

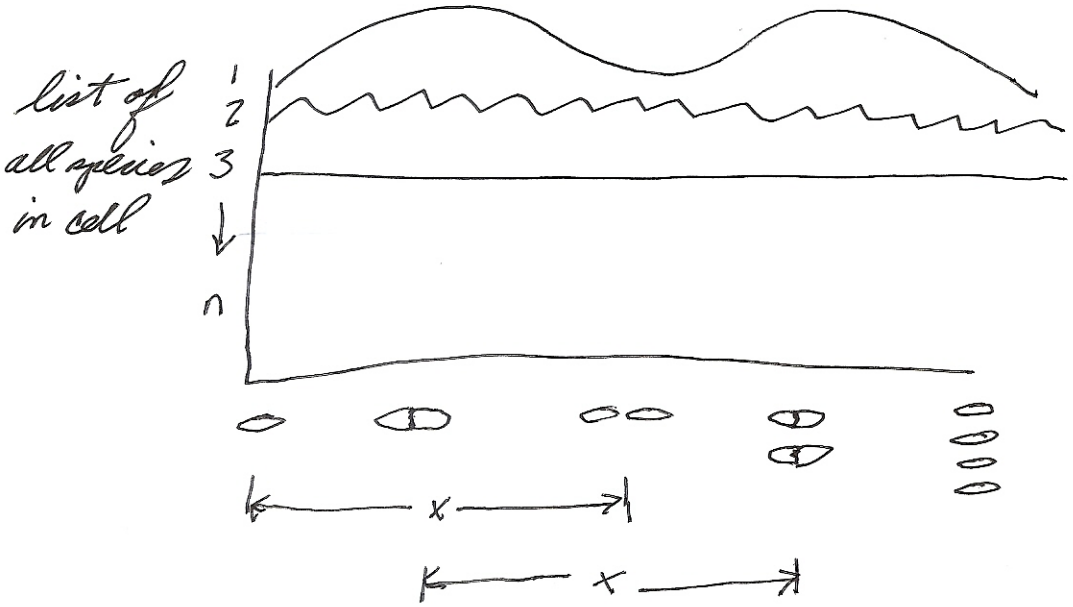
13. List the regulatory factors involved in the nucleic acid mainframe?

    1) ds sequence plus monomer concentration → distribution of RNA via the 2°/3° conformational dynamics of the ds sequence/monomer interaction.
    2) 2°/3° sequential folding of RNA determines its fate in the cytoplasm and # for message.

    So the proteins and stable RNA made are primarily determined by the kinetics of nucleic acid conformational dynamics utilizing the monomers as the informational input.

14. What is the cell cycle?

    The cell cycle is the change in concentration of the various species of the cell such that the binary fission cycle is obtained. A binary fission cycle implies that over all measurements one period separated or one cycle separated each species id doubled, and if measured from the point of daughter separation there are two cells from one.



Over all periods of x the concentration per "cell" is constant.

15. How and what meaning is to be attached to the cyclic compositional behavior?

We want to relate the cycle for each species to each other. This is the regulatory cycle of the cell. At constant growth conditions it is a central notion that the cyclic behavior is not just a simple constant program being played out over and over or read off the DNA molecule but rather is a cycle being continuously operated through interaction among the species, at the deepest level or the root of the architecture being determined by the continuous evolution of the monomer/sequence interaction. This is a subtle point and means that even at constant growth conditions the regulatory architecture is effecting adaptive response. The same regulatory architecture is being used at constant growth conditions in different environments as during transitions, either ups or shifts downs. During a transition the individual species of the cell will not be constant over one period of the cell cycle but will be different.

16. What then is the regulatory architecture in the cell?

    1) The fast time scale regulation in the cytoplasm effected by allostery and cooperativity, oligomer and complex formation built into proteins, effecting catalysis which regulate the small molecule precursor synthesis from the mass/energy exchange over the cell boundary, essentially a kinetic net with fast throughput of small molecules.
    2) The monomer → macromolecular (nucleic acid and protein) regulatory system determined by the kinetics and conformation dynamics in the nucleic acid mainframe.

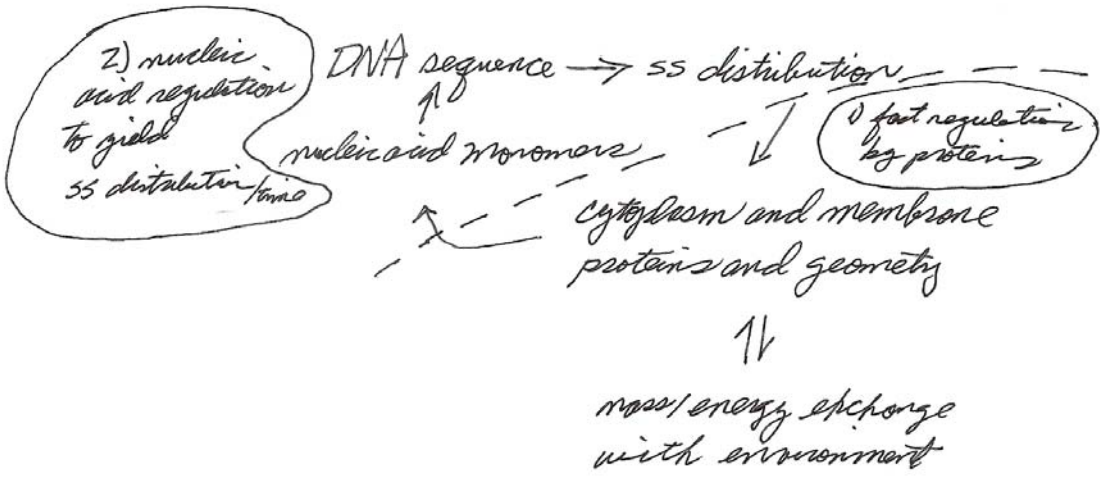
    The macromolecular regulatory system is completely surrounded by the cytoplasmic regulatory system within the cell, information and mass transport being defined by the continuous vectorial flow between the two systems across the dynamic RNA processing interfacial area, this flow being a dissipative structure composed almost entirely of nucleic acids and the support proteins, RNA polymerases, histone like proteins, etc.

17. Reevaluate the Central Dogma?

    The Central Dogma is incomplete. It must be a cycle and not linear. The information in the surrounding environment, the composition of the environment, must be the primary input. Given the geometry of the cell, the transport and diffusivity of molecules, ions and H_2O into and out of the cell, and given the proteins possible from the DNA molecule, population dynamics will always lead to enrichment of the fastest growing cells. Fastest growing means that the most monomers for macromolecular, membrane and cell wall synthesis are produced, of course observing the stoichiometric balance required for cyclic evolution of the species through the cell cycle and this balanced small molecule synthesis will imply a certain stoichiometry of mass/energy exchange across the cell boundary through time.

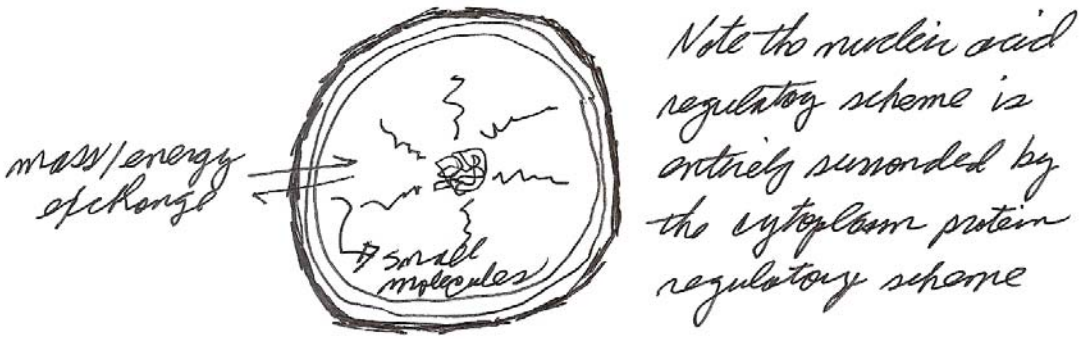
    The Central Dogma can be expected to accommodate communication with the environment, the thermal code and the nucleic acid mainframe. The arrows imply

physical transport of the respective chemical species, transport being completely understood in terms of the phases and the dissipative structure.



The mass/energy exchange provides continuous addition of material and the Central Dogma in cyclic form with this addition, closes over the cell cycle.

A cartoon of the spatial Central Dogma would be:



18. What desirable properties due to symmetry does the Nucleic Acid Mainframe have?

1) The sum of all momentum of all species involved in the flow system is ~0.
2) ss products are distributed in a random fashion over the surface of the nucleoid.
3) Separation of monomers from the cytoplasm is carried out equally over the entire nucleoid surface.

The ∑momentum = ~0 is required by simple conservation equations. Without this the mainframe would translate. This will be inspected more closely when chromosome segregation is considered. The sequential nature of ribosome

maturation, ribosome binding to messenger and folding or 2°/3° structure formation of RNA requires the random distribution of ss to the nucleoid/cytoplasm interface. Since these are kinetic processes, all ss must be treated similarly or the regulation through competition of nucleases, ribosomes and regulatory proteins for sequential structures is lost.

So all the nice properties of symmetry respecting the mainframe are related to transport. We want to minimize transport work and maximize orderly processing of information.

19. What is the total work of making ss and how is this related to the minimum entropy production characteristic of dissipative structures?

    The work is (all these are already defined spatially):

    1) Separation of monomers from the cytoplasm.
    2) Transport of monomers to polymerization centers.
    3) Transport of ss away from ds.
    4) Processing of ss.
    5) Getting RNA polymerase to the right spot.

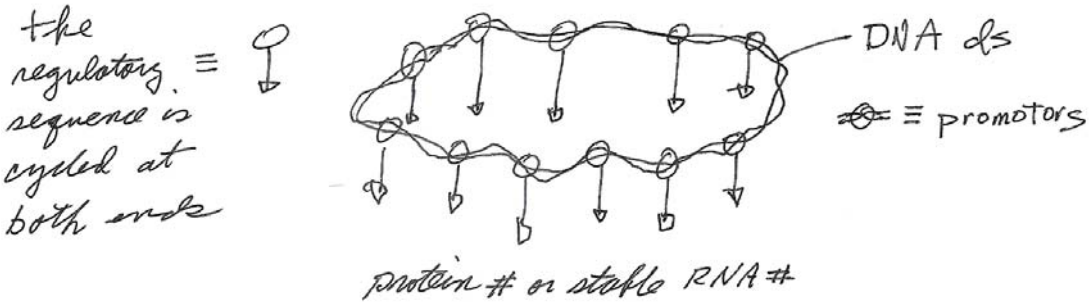
    Minimum entropy production in the microscopic world may be roughly construed to mean the shortest transport paths for all species, separation work spread over the largest interfacial surface area allowing the lowest density of separation work per unit area meaning closest to equilibrium, the maximal utilization of the irreversible input to the system by relaxation following electronic transitions (polymerization reactions) and looking as close as possible to the random walk of species which characterizes equilibrium (equilibrium is of course 0 entropy production). In a system involving polymers whose 1° (primary) structure is restrained from obtaining 2°/3° structure, such as with ss in the nucleoid, it should be observed that the ss is almost immediately in thermal equilibrium by relaxation from the polymerization reaction and utilizing the difference between equilibrium configuration of the 1° in the nucleoid and in the cytoplasm to drive the vectorial transport, and undergoing a random walk to the interface to be able to obtain this, represents the maximal use of chemical bond energy released in the polymerization reaction that could be imagined. For instance, if the DNA was dispersed and RNA could fold as it is synthesized without obtaining any transport, this would simply amount unutilized thermal dissipation. If the reaction centers in the nucleoid are observed through time they would appear to be randomly distributed throughout the nucleoid. This arises due to mixing and statistical parameters of the topology of the DNA in the condensed nucleoid and its transcriptional activity and gives the equilibrium appearing distribution, the minimum entropy production configuration. It should be noted that this level of microscopic thermodynamic analysis is here a subtle and creative business and quantification of this system will probably allow a deeper level of comprehension of entropy then is currently possible by

physiochemical approaches. The true nature and meaning of entropy is available by quantification of this system.

The work associated with processing of RNA brings a somewhat different concept. It will be desired that the task is obtained with the smallest number of proteins, ribosomes and nucleases. This can be effected when these molecules are confined spatially and recycled, as they are in the RNA processing zone.

20. Discuss the relation between equilibrium and the regulatory structure?

First we will take an overview of the process. The spatial organization or where in the cell these processes go on has already been discussed and what motivates the flow.



The nucleic acid regulatory structure spans from the DNA sequence to the final protein or stable RNA product. The process will be reviewed in three different steps and then integrated:

1) The equilibrium seeking phenomenon relevant to each level of the regulatory architecture.
2) Perturbations from equilibrium which motivate these and allow the sequential processes to become cyclic.
3) Sequence specific recognition steps within the sequence of conformational or transport equilibrium seeking events.

The first has already been discussed and consists of the following processes:

1) Diffusion of monomers into the nucleoid at stationary state or multiple steady state flow of the mainframe. This is a small molecule transport seeking equilibrium.
2) The Thermal Code selective destabilization of the DNA ds represents the DNA sequence seeking its equilibrium configuration given the sequence and monomer composition data assuming the other factors that are in the background such as ionic strength, average superhelical density, polyamine, HLP, concentrations, etc, only provide the environment where the sequence and monomer data evolve the signals which are utilized.

3) The ss migration out. Interesting in that this builds in a time lag, the random walk to the surface of the nucleoid, obtaining equilibrium rapidly both in the cytoplasm and the nucleoid, and the connectivity between this lag the RNA defines the vectorial flow.
4) ss 2°/3° formation is the folding of the polymer seeking equilibrium, folding as it sequentially emerges from the nucleoid.
5) → n) Mass/energy exchange to monomer production in the cytoplasm with transport over the boundary.

n + 1) With the monomers generated or recycled we go back to 1). This equilibrium seeking sequence becomes cyclic.

The perturbations which fit into these sequential equilibrium seeking processes and allow the transcription initiation and translation initiation to become cyclic events.

1) Monomers are made in the cytoplasm leading to a monomer equilibrium configuration resulting in transport, both from a stationary phase or when the mainframe is running, allowing it to keep running.
2) RNA polymerase binding captures the destabilization structure formed and translocation accompanying propagation returns the DNA to a closed state where it can again evolve to a destabilized structure, cycling the transcription initiation event. (kinetics determined by the Thermal Code)
3) The continuous addition of ss within the nucleoid associated with propagation provides ss necessary for the vectorial flow.
4) Interaction of ss 2°/3° with ribosomes, nucleases and proteins perturb the equilibrium directed folding of the ss and translocation associated with polymerization allows the 2°/3° to again interact, allowing cycling of the translation initiation process. The kinetics of this 2°/3° equilibrium structure formation and its perturbation through initiation and reformation with translocation determine the translation # for each message and define the equilibrium seeking/perturbation kinetic cycle for message.
5) → n) Mass exchange to monomer production in the cytoplasm.

n + 1) Back to step 1), the monomer non-equilibrium configuration which motivates net transport.

The sequence specific recognition events which fit into this sequential cyclic regulatory architecture and insure absolute fidelity by giving to the regulatory architecture equilibrium sequence “orthogonal” equilibriums (orthogonal in that the other equilibrium steps deal with conformations of polymers or configurations of small molecules in space). These recognition events are embedded in the dynamic sequential nucleic acid structures already outlined. The protein or nucleic acids (ribosomes) with which sequence specific interaction recognition events occur come to a “local” equilibrium with the specific sequence to effect recognition. They are in the sequence regulation equilibrium seeking events such that preceding structures in the regulatory sequence must come first and then after the recognition event succeeding steps are allowed to occur.

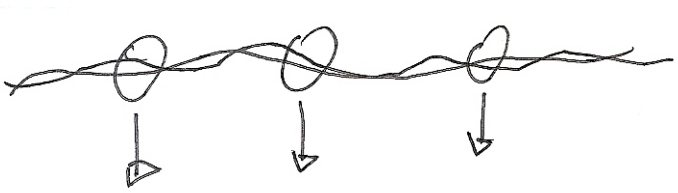
1) Upstream and Pribnow (or whatever sequence is recognized by other than σ protein of the holoenzyme, δ, δ', etc) define possibly (probably) strand selection and the Pribnow (or analogous sequence with other holoenzyme) define a clean start site – the exact base at which transcription will begin. So the transcription initiation event requires first DNA structure formation (Thermal Code), then RNA polymerase binding, then the specific recognition between polymerase or a polymerase subunit, a sequence where bound on DNA, and then is ready to go with a clean start site defined on the DNA. As mentioned earlier, the rate of polymerization from this complex may have 16 possible rates if the first two bases are somewhat special – an end effect. At this point we have extreme high fidelity for transcription initiation (the fluctuation of monomer concentration and DNA conformational smoothing to produce high fidelity or precisely defined kinetic parameters for the destabilization event have not been discussed yet, and can only be discussed in terms of coupling with the collision rate and fluctuations between DNA and RNA polymerase.
2) Shine – Delgarno. Here presumably first there is a course recognition of some 2°/3° in the message in which the S – D sequence is embedded. This then leads to the start sequence codon recognition. I don't understand recognition dynamics for this process, but clearly some sequential combination of structure and sequence specific events occurs so that the kinetics of these events determine high fidelity and copy # regulation.

This sequence of events motivate the fidelity and regulatory kinetic control. These are 2°/3° equilibrium seeking nucleic acid phenomenon coupled in a sequential pathway with sequence specific recognition of sequences in these dynamic structures (orthogonal controls). All this regulatory sequence is available at an early evolutionary stage because it's all in 1° sequence of nucleic acids with only a small number of "nice" proteins needed.

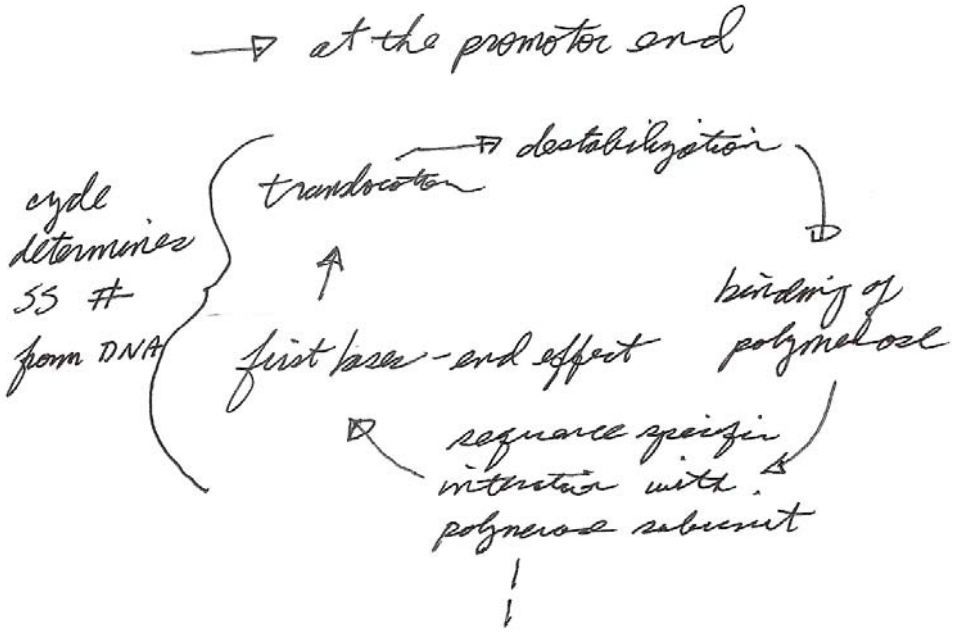
So the sequential hierarchy looks like:

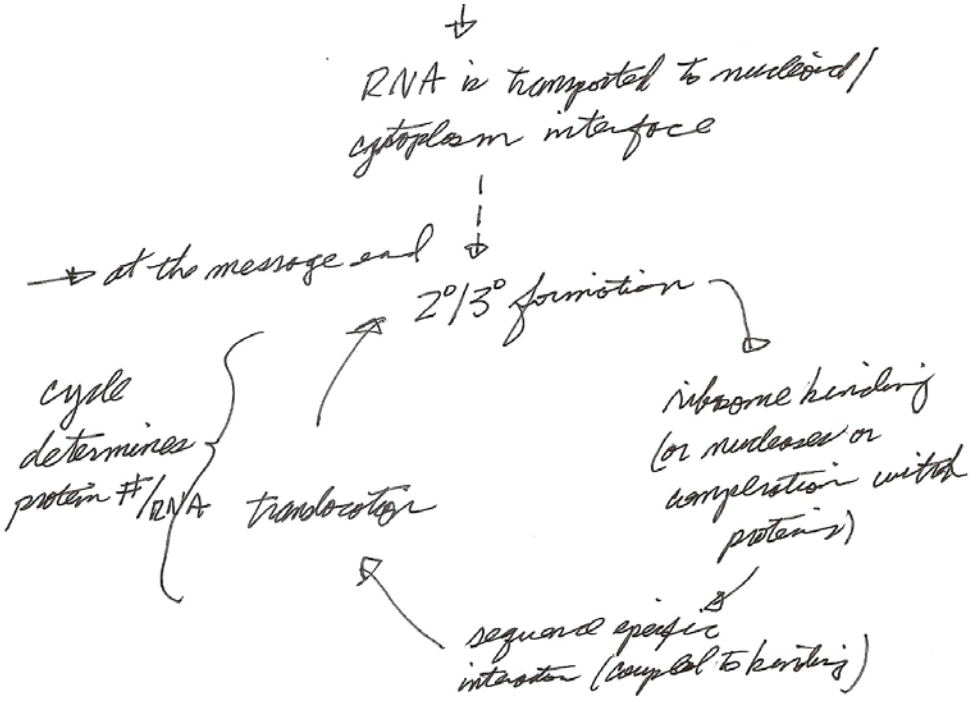
1. destabilize ds
2. bind RNA polymerase or allow in
3. sequence specific recognition for clean start and fidelity of polymerase with destabilized zone
4. rate modulation due to end effect – first one or two phosphodiester bond formation rate dependent on end di or trinucleatides.
5. 2°/3° folding of RNA
6. interaction of RNA with ribosomes, nucleases, or proteins (coupled or not with 6)
7. Shine Delgarno recognition

and a sequence like this proceeds from each promotor on the DNA molecule



and the distribution of the primary events, DNA ds destabilization is determined by the Thermal Code and perturbation or capturing of structures in the sequential pathway by RNA polymerase, ribosomes, nucleases, proteins and translocation lead to a cycling of the regulatory sequence at both ends





At both ends, neither ribosomes or polymerases can limit the reaction sequence, it must rather be limited by the nucleic acid structural substrate formation or control through control through these mechanisms cannot be effected.

The information of the cell "condition" is fed in at the DNA destabilization step – the Thermal Code, RNA 2°/3° appears not to have an environmental input except through competition between nucleases/ribosomes/and regulatory proteins.

The sequential nature of the regulatory pathway from the primary event or root of the regulatory architecture, the DNA destabilization, strongly implicates, if not tautologically implicates that the majority of the control of the cell cycle (changing distribution of proteins through time) must be at this step (to say nothing of the fact that only the Thermal Code has anywhere near the information processing capacity for the job of changing the distribution of proteins through time)

21. What is the striking feature(s) of replication initiation?

    1) It is the only thing in the cell that happens exactly once! (neglecting those processes which also happen only once but are in a sequence whose root is replication initiation) Given the fluctuations of the thermal environment this is no small task. And it always happens at least once (in a growing culture). A very precise event.
    2) The replication initiation event permits complete replication of the chromosome (come hell or high water) This unequivocally requires a forecast of the amount of monomer needed to complete the task and therefore it must be possible to synthesize that quantity of monomers

22. What fundamental physical principle and abstract abilities are implied by these two features of the replication initiation process?

    That it always happens once and only once (ignoring for the moment re-initiation leading > 2 genomes in a cell) implies that the signal is an equilibrium motivated or seeking signal such that if it does not occur when the information input says it should it will occur just a little later, but as we get further and further from the proper point of initiation the probability that initiation has not occurred converges to 0. This is just the notion that equilibrium or an equilibrium seeking switch or signal (a structure or conformation) cannot be kept from seeking equilibrium in the absence of a kinetic or mechanistic barrier unless the composition changes such that the equilibrium is no longer defined.

    That given an initiation event replication can proceed to termination or completion implies that an absolute quantification is made, an inventory assessment, which must be supported by a counting or number system. This is the abstract part. A counting system is essentially required. There is no getting around this.

23. What molecular mechanism can provide these required properties?

    The Thermal Code. The destabilization of the DNA ds is an equilibrium seeking event which is a function of sequence and monomer concentration. If the sequence and monomer concentration require that a given stretch of DNA ds is destabilized it will obtain unless it is blocked by another species (a nice protein for a virus to make perhaps)

    The absolute counting requirement sheds light on the computer nature of this entire construct and helps to motivate the Thermal Code in both a molecular and abstract sense.

    Any computer requires some counting system, which does not have to be a linear one. The Thermal Code provides a non-linear counting system. Since the conformation or destabilization that arises in the DNA ds represents a sum of sequence specific interactions over some finite distance of the ds array of bases and since the sum is a function of monomer concentration in the nucleoid around the DNA ds a counting system is effected. Viewing the ds as an array with the monomer interactions being a function of the monomer concentration provides a counting system in a simple mechanistic way, although the non-linear properties make the summation or relation between the DNA conformational dynamics somewhat difficult to get out.

    The Thermal Code replication initiation signal can determine absolute quantities because the DNA ds is of finite length, it offers a certain number of liganding sites,

so by relating the total number of liganding sites to the monomer concentration, a sequence of DNA can be defined that does a conformational transition in relation to the absolute monomer concentration in the cell. This is a subtle point, perhaps obscured in this explanation, but very important. The relation between absolute counting and relative counting will be expanded on later. A Thermal Code signal which is dynamic, arising many times throughout the cell cycle, does not have to make an absolute assessment of the monomer concentration, but the replication ori destabilization which arises only once in the cell cycle and by virtue of the finite # of liganding sites on the DNA molecule can make an absolute determination of monomer concentration in the cell.

It should be noted that this absolute/relative counting distinction is a programming specification, i.e. a sequence of DNA can be constructed with both properties and for most promotor sequences the relative opening rate is what is critically important to the regulation, a global assessment of monomer concentration being implicit, while for the replication ori, it is the global assessment that is desired.

24. Discuss multiple replication initiation through the cell cycle and how this relates the absolute/relative programming specification. (programming always refers to the 1° sequence of the DNA) Relate this to the evolution of the monomer concentration and the equilibrium seeking nature of the Thermal Code.

    Replication initiation is a very slow event. It may take a long time before the absolute concentration requirement is met for ori destabilization. Once this concentration is met, the probability of the ori opening converges to one. This equilibrium seeking conformational transition motivates the fidelity for doing something once and at least once. To prevent re-initiation requires that the ori destabilization leads to a sequence of events which modify the monomer concentration, or bring it back down below the critical absolute concentration which was the initiation information input. The replication process provides this mechanism by siphoning of monomers from the NTP pool at a rapid rate, bringing the concentration back below initiation conditions.

    After replication has proceeded for awhile, the monomer concentration may evolve back up to the initiation condition, only this time there are two ori sequences and then both of these will open or re-initiate replication.

    This points out the nature of an absolute concentration determination via the Thermal Code. If a sequence of DNA is destabilized above some critical concentration of monomers and the monomer concentration remains above that critical concentration then it will reopen over and over again, by the equilibrium seeking/perturbation (capture of structure) cycle and the process becomes dynamic with some kinetic rate for this cycle. Once the cycle is dominated by the kinetics of the process, absolute concentration determination may or probably is lost, all that can be said is it is over some critical level with Thermal Code interactions determining the rate to the extent possible, but if the monomer concentration

evolves back below the critical opening concentration, then the rate is determined by the concentration evolution back to that critical or absolute level and not by the kinetics and dynamics of the cycle. This is precisely the nature required for the replication initiation decision, the special sequence of events occurring after ori destabilization allowing the replication initiation to be an absolute concentration determination event.

25. What does the replication initiation or other very slow destabilizing phenomenon say concerning the lyotropic vs thermotropic nature of the thermal code?

    These slow destabilization events would indicate that the structure of the DNA is highly dependent on the monomer concentration and the simple sequence dependent thermotropic behavior is not so important.

    The fine concentration measurement required for replication initiation also strongly suggests this.

26. At the other end of the scale, promoters for ribosomal RNA, what is implied there?

    Looking at promotors which are as fast as possible will indicate the kinetics of the dynamic equilibrium seeking destabilization/perturbation through capture cycle. Just how fats can the cycle be turned over? Rough calculation on the stable RNA promotors indicate reinitiating every 1 – 1.5 seconds would satisfy known production rates. That's per promotor and the ribosomal RNA has tandem promotors so some cooperativity must be looked at.

    To investigate the kinetic parameters requires first a look at the entire molecular mechanism involved with transcription, not just looking at an individual segment of DNA.

27. Since the migration of ss out of the nucleoid from the point of polymerization clearly requires the RNA to not be wrapping around the helix as it is polymerized, i.e. it has a fixed average geometry (vectorial flow) relative to the nucleoid, and the strand being copied is in a helix; there is a topological problem. How is this resolved?

    The DNA must twist through the fixed polymerase/ss complex.

28. How can this occur without distorting the destabilization/twist/coiling (coiling ≡ supercoiling) balance which must be so sensitively balanced for the Thermal Code to work?

    The coils are in a dynamic equilibrium throughout the nucleoid. They exchange and migrate. There is both a short range equilibrium over a stretch of DNA and long range equilibrium between segments of DNA far apart by backbone connectivity

but close due to the condensation. This leads to a global twist/coil equilibrium throughout the nucleoid.

By collapse of a coil ahead of the transcribing polymerase complex and formation of a coil behind the complex, twisting through the fixed geometry polymerase complex is achieved.

29. Wouldn't this lead to a short range non-equilibrium configuration along a segment of the DNA?

    The long range exchange of coils between adjacent topologically but distant by connectivity alleviates this. The global equilibrium in the nucleoid continuously relieves the local non-equilibrium configuration.

30. What does this imply about the coil forming process?

    It is very reversible, there should be thermally motivated exchange, the Boltzmann distribution of coils is populated at room temperature. These properties of the coiling phenomenon are well known.

31. What if the coils are stabilized more strongly by interaction with some protein species?

    The migration of coils would be restricted.

32. What species probably stabilize coils?

    Polyamines and HLP's present a relatively rigid electrostatic topology which can interact with some repeating electrostatic topology found in the coiled DNA, thereby stabilizing these structures. But the degree of stabilization must be carefully set so as to obtain the desired equilibrium twist/coil level and migration rate.

33. The exchange of coils then involves losing or acquiring polyamines or HLP's?

    Yes. These can exchange readily in the condensed DNA because of the adjacency but could not with the DNA dispersed.

34. How much of the DNA is in coils in vivo?

    Coil density, σ, is found to be ~.05 – .06. This would lead to $2 – 2.4 \times 10^4$ coils/genome equivalent in E. coli. At a coil size of 180 Å in diameter this means practically all the DNA is coiled.

35. Summarize then the background in the nucleoid for the Thermal Code?

Almost all the DNA is coiled. Coils are in short range equilibrium with twists – grab a section of DNA and twist it and one way puts in coils, the other way takes them out. Coils are in long range equilibrium by propagation of this short range equilibrium over the connectivity of the backbone, this cannot propagate beyond the persistence length (~600 Å) by definition and long range or global equilibrium is obtained by exchange of adjacent short range segments in the condensed structure. These exchange phenomenon are quite reversible, thermally populated. The dynamic equilibrium defined by these processes establishes the background on which the Thermal Code works. The Thermal Code model then asserts that the configuration of destabilized areas on the DNA ds will be determined by sequence and monomer concentration, σ and coil ligands establishing and restoring and maintaining the DNA such that the signal channel or determinant of the equilibrium configuration is due to the DNA sequence and its interaction with monomers.

36. Discuss the Thermal Code in terms of this equilibrium configuration, reversibility, and balance?

    The reversibility of the decision making process of where access memory is the most important aspect of this entire model. Reversible or near reversible access to memory is the central requirement in a computer or computer like machine. Reversible means close to equilibrium, very small entropy production. This is precisely what obtains in the nucleoid. The destabilization/twist/coil is finely balanced by the previously mentioned background phenomenon, with the distribution of destabilized areas being determined through the signal channel of the Thermal Code, the destabilization being an equilibrium seeking process between the sequence and the monomers, leading to localization of destabilized areas. Given the background, sequence and monomer concentrations a certain global quantity of destabilization must exist, the question is where and the Thermal Code determines where. It can be viewed as a wave localization. The localization of these destabilizations becoming cyclic through the perturbation or capture by polymerase (gets into helix) and sequential events from this primary one as already discussed. The balance is determined by the DNA connectivity, its equivalence throughout the nucleoid, all being treated equally and identical except for the conformations that obtain as a function of sequence at constant monomer concentration and the distribution being altered as monomer concentrations change. This becomes motivated at a very fundamental informational level since the sequence is the only information in a system at constant environmental concentration and the change in monomer concentration through time is the primary informational input. It should be remembered that the abstract potential for information in the monomer concentration is ∞ and the number of liganding sites which are used to determine a conformation to produce a recognizable physical signal.

37. How are fluctuations and gradient problems in the microscopic environment of the nucleoid handled?

The global mixing, equivalence of DNA, throughout the nucleoid and high diffusivity of monomers in the nucleoid keep gradients from forming in the majority of the volume of the nucleoid. Clearly there will be a gradient and much commotion around the interface with the cytoplasm where the mass exchange is occurring. Gradients within the nucleoid will be unstable because they represent a non-equilibrium configuration. The equilibrium configuration and definition of a phase continuity of constant composition, net of thermal fluctuations. The mixing (this refers to DNA undergoing translation and rotation) and diffusivity prevent and minimize kinetic barriers to this equilibrium configuration.

The microscopic fluctuations are smoothed by the Thermal Code mechanism. It represents a sum of interactions possibly up to the persistence length of the DNA. This smoothes microscopic fluctuations in the concentration space very nicely. This smoothing is at the heart of the fidelity of information transfer between the DNA molecule and the environment. It is what both smoothes local fluctuations and allows fine concentration resolution. The smallest difference in interactions between different segments of the DNA sequence can lead to the choice of which segment is destabilized. They are competing to localize destabilization, the global amount of destabilization being fixed, so are virtually indistinguishable energetically. If the DNA sequence was composed of some constant repeat, the list of destabilized areas would be energetically degenerate. Two promotors that had different sequences but the same destabilization pattern would be degenerate energetically. It is this closeness to degeneracy in energy for different patterns of destabilization which allows the abstract power of the Thermal Code (in terms of concentration space for monomers and long array for the DNA) to be realized.

38. Motivate clearly the thermotropic (sequence dependent only) vs the lyotropic (sequence and interactions with monomers) distinction?

    When the nucleic acid mainframe is running at steady state and the concentration of monomers is in steady state flow the pattern of RNA ss coming out of the nucleoid can only be a function of the DNA sequence. This is the thermotropic or strictly sequence dependent behavior.

    Note that steady state will be exceedingly unstable since the environment in the cytoplasm is continuously changing. geometrical changes in the cell alone will lead to evolution of composition. This steady state is dramatically opposed to the binary fission cycle (it is what we will want to program into cells for chemical transformation technology).

    If we hold the sequence and background support of the nucleoid constant and vary the monomer concentration the pattern of RNA coming out will change. This is the lyotropic behavior. This is almost always the situation in the cell.

Of course in the cell sequence changes through time also when replication is proceeding. In the absence of this lyotropic response it would be impossible to ever get to replication initiation.

Quantifying the relationship between global RNA synthesis at different compositions but a similar summed concentration of monomers will be very interesting. Of course this will have to be tightly coupled with the monomer production in the cytoplasm.

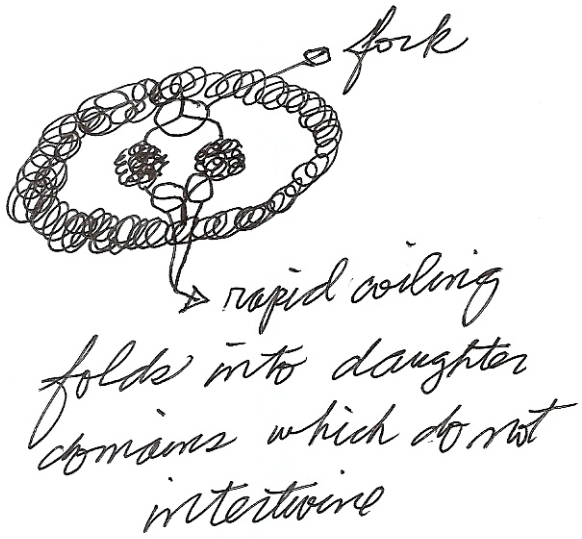
39. So the thermotropic and lyotropic are not in conflict but really orthogonal?

Yes. At constant composition the equilibrium seeking destabilization formation/perturbation through capture cycle will turnover at a different rate dependent on sequence changing the composition and will change these rates. The two ends of the spectrum may be represented by ribosomal promotors and the replication initiation promotor. Ribosomal promotors turn over or cycle rapidly and are relatively insensitive to monomer concentration whereas the ori is exceedingly sensitive to monomer concentration. This represents a difference of approximately three orders of magnitude. This is the maximal extent of regulation possible through the Thermal Code or by any mechanism at the level of transcription since initiation every 1 – 1.8 seconds is the wide open limit and initiation once per cell cycle the minimum. The degree of variation possible for destabilization formation over the concentration range which occurs in vivo in the nucleoid is an exceedingly interesting parameter. To what extent and the nature of the interaction between lyotropic and thermotropic will give good quantitative data.

40. Describe the DNA replication process in terms of equilibrium and topology?

The replication forks meander through the nucleoid such that if it was observed through time it would appear to be a random walk. This is the equilibrium configuration respecting DNA ds transport – there is none. Diffusion of monomers to the forks and separation from the cytoplasm (assuming they are not synthesized in the nucleoid) are done with the minimum entropy production by the nucleic acid mainframe.

Rapid relaxation to a coiled configuration behind the forks prevents intertwining of the newly replicated strands. These are provided by exchange of coils throughout the global equilibrium of twist/coil in the nucleoid. These coiled structures do not intertwine or mix because this is topologically forbidden for compacted, coiled DNA. Addition of supercoils to the nucleoid is required to maintain the coiling density since one ds is going to two and all is coiled requiring net addition.

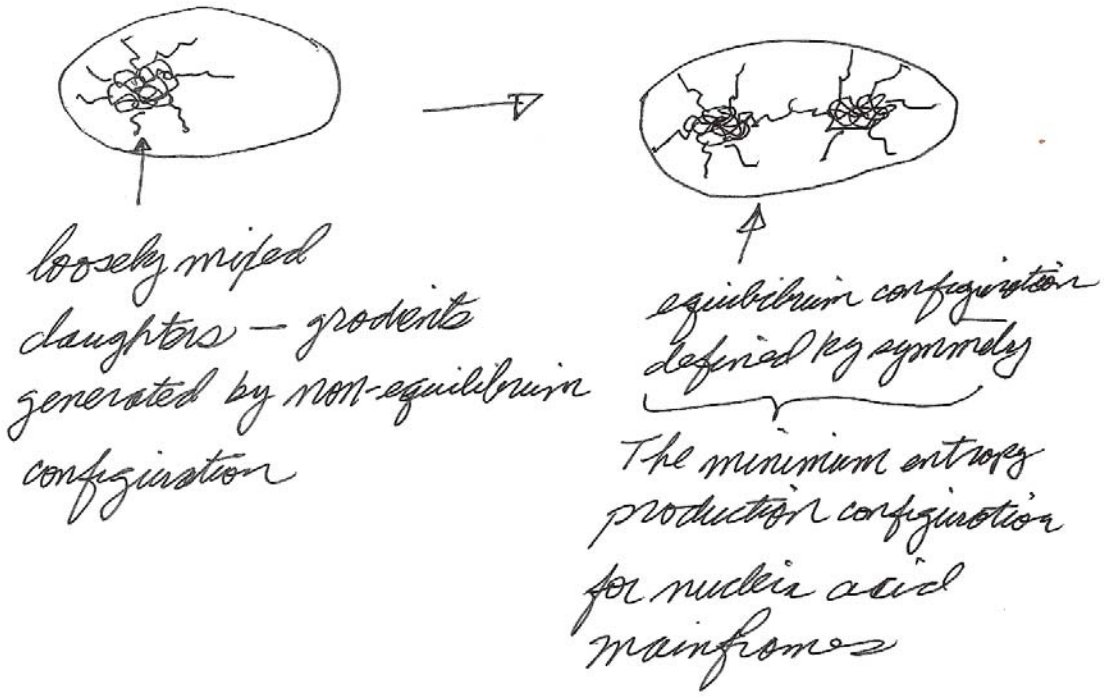


The domains of coiled DNA behind the forks, the daughter nucleoids, grow and mix in a gross way such that when replication is complete, although the domains may be intermixed, the ds themselves are not.

41. How do the replicated, domain mixed, daughter DNA molecules segregate after completion of replication?

    This is a spontaneous equilibrium seeking process. The nucleoid symmetry in the cell is defined by equilibrium seeking configurations of the nucleic acid mainframe. If it migrates to one corner of the cell, monomer and ss gradients will build up and provide a restoring force back to the center of the cell. Only when the nucleoid is in the center of the cell is transport minimized, so the symmetry defines this as the equilibrium configuration.

    After replication into the loosely mixed daughter domain structure, this equilibrium configuration seeking process will lead to a migration apart of the daughter nucleoids.



So connectivity is required for maintenance of the nucleoid phase.

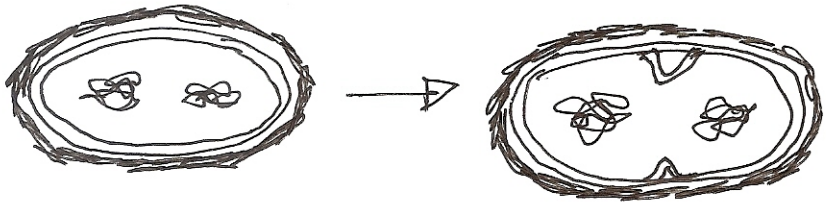
42. What is origin or root of asymmetry in the binary fission cycle?

    The spontaneous or equilibrium seeking segregation of the replicated DNA.

    So the origin of asymmetry has a simple motivation due to the nature of the nucleoid phase and its flow system.

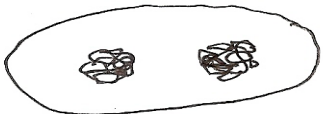
43. If this is the root of asymmetry, does that imply the separation process organization follows or is brought about by this primary asymmetric event?

    Yes. By altering the rates of cell membrane and cell wall synthesis such that there is an excess of cell membrane synthesis bulges will spontaneously appear in the cell membrane, again symmetry defined by an equilibrium seeking process, at the center of the cell between the two nucleic acid mainframes.

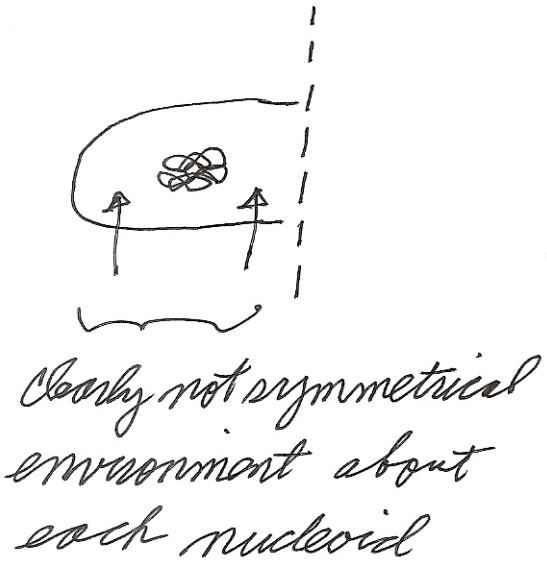


    Insertion of cell wall material into the cell membrane bulges leads to cell wall growth in the forming septum (this is superficial, but the motivation is fundamental,

the only point being the septation process can be motivated by the nucleoid segregation).

44. This configuration,  , appears symmetrical. Why call segregation the origin or root of asymmetry?

    The configuration is still symmetry dominated and completely symmetrical except for one element. Each nucleoid is now in an asymmetric enclosure.



45. Can the asymmetry be taken back to an earlier step?

    No. The DNA molecule is itself replicated by the two growing forks growing from the replication ori and the symmetry of the helix itself is obvious.

46. Go back to the thermotropic/lyotropic discussion respecting helix destabilization and examine the circularity.

    We state that at the steady state operation of the nucleic acid mainframe the distribution of RNA put out must be a function of sequence, since the monomer concentration is constant. However the turnover of the destabilization cycle involves interaction between monomers and the DNA sequence and this is why when the monomer concentration changes, the destabilization cycle changes, what we are calling the lyotropic response.

    Clearly we have a circularity problem here. The Thermal Code will have to be completely elucidated before this problem is resolved. The Thermal Code is obviously completely general, stating only that the destabilization pattern is a

function of sequence and monomers in the in vivo background already described which allows sequence and monomer concentration to be the determinants of the pattern of structures that evolve. By general is simply meant that one can have any sequence and any monomer concentration, from 0 → saturation limit for each possible monomer.

47. Do in vitro promoter specific binding studies represent a problem for the Thermal Code construct, since correct RNA polymerase/promoter complex formation occurs in the absence of any monomers?

    The Thermal Code states the recognition event is a destabilization. The DNA stability in these in vitro systems is such that the fluctuational events along the DNA helix can provide a primary destabilization signal which RNA polymerase can interact all along the ds and the 2° sequence recognition event with Pribnow or other homology leads to the stable complex at the promotor.

    So the presumed sequence of events for transcription initiation by the Thermal Code construct can explain the binding studies.

    The generally AT rich nature of promotors would also localize DNA helix fluctuations in these areas.

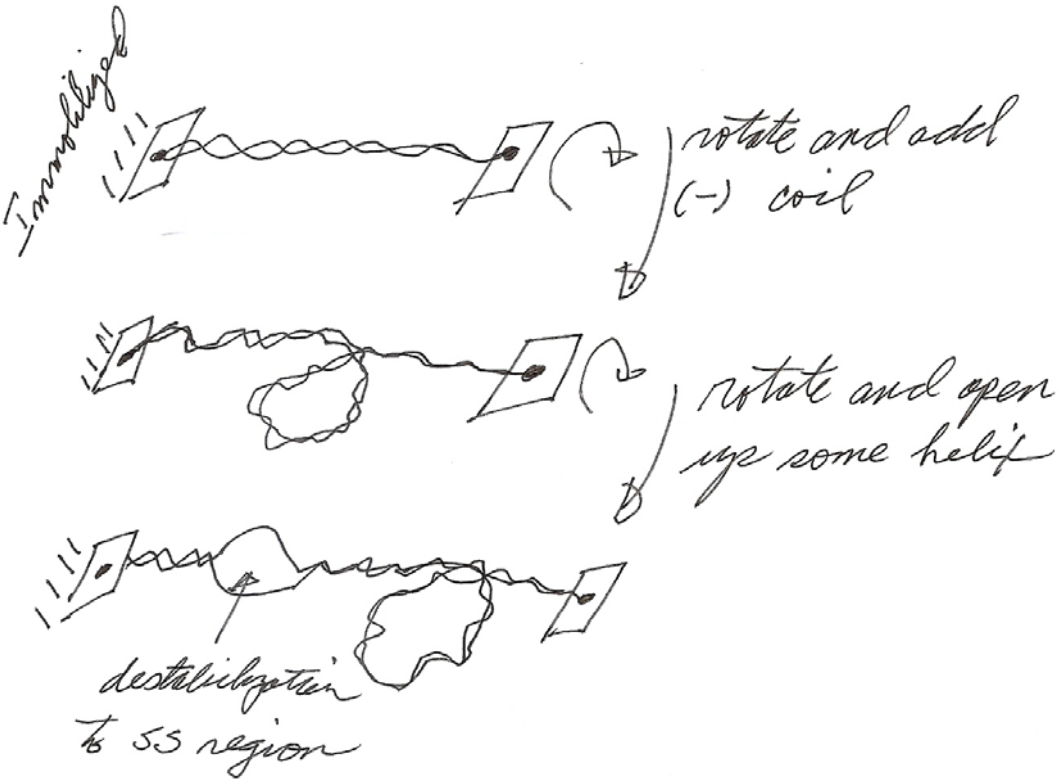
    The incredible balance between near equilibrium for the DNA helix stability which is finely dissected by the Thermal Code appears to be largely unavailable in these in vitro systems so it is possible to rationalize the results of the in vitro binding studies but not to prove, disprove or elucidate the kinetic parameters and construct of the Thermal Code.

48. Discuss the global supercoil/twist/Thermal code dynamics in terms of torsional stress equilibrium.

    The concept of torsional stress is central to understanding the DNA helix stability, topology and the dynamic behavior, both at equilibrium fluctuational dynamics and leading to non-equilibrium complexes and the equilibrium seeking cycle these provide a mechanism for.

    The notion is that torsional stress is in equilibrium over the connectivity of the DNA molecule in the nucleoid. This means it is constant, net of thermal fluctuations, over the molecule. The different conformation and topology of the DNA molecule reflects the fact that the expression of torsional tension or stress is different for different sequences. The stress is constant, the conformation is different. All interactions are included in this. Differential interactions dependent on sequence or conformation are included. The goal is to finally understand how sequence and environment composition factors lead to different conformations over the ds at a constant average level of torsional stress over the ds.

To state in a different way:



In all three situations the DNA helix will have an equilibrium pattern of conformation and topology.

The average torsional stress is different in each because we have rotated one end and then holding it fixed.

To say the torsional stress is constant over the DNA molecule in each of these situations is to say that in a plane bisecting any adjacent basepairs, irrespective of their being in a coiled section, destabilized section, or whatever conformation, the time average (to remove thermal fluctuations) torsional stress or through that plane is constant. The different conformation then reflect differences in sequence and differential interactions with the environment composition.

A requirement for this to be true is that all conformations can obtain equilibrium through thermal processes; there are no barriers between conformations too high for thermal fluctuations to provide a transition pathway.

Long-range torsional stress equilibrium over the ds obtains rapidly in the nucleoid by exchange of supercoils and coil stabilizing ligands between adjacent DNA strands in the nucleoid. Otherwise it would take a very long time if local stress had to be relieved by following connectivity. The persistence length of in vivo DNA ds as measured by stressing local areas and observing how far away these stresses propagate is strongly related to this short-range, long-range, equilibrium of

coil/twist/Thermal Code equilibrium seeking mechanism. The importance of condensation of the DNA is rather clear.

All of this is just a qualitative description and quantifying these phenomenon will be required to understand the in vivo operation of the DNA molecule. Quantification of phenomenon like these is an art strongly relying on the theory of computation, first principles, understanding the microscopic environment, sifting out a tractable view while retaining information, getting a description with utility, etc, so it is impossible to present how I really feel and view this quantification problem.

49. Discuss thermal adaptation in reference to a regulatory mechanism.

    Cells grow at different rates in mediums with a different composition and at different rates at different temperatures. Given the definition of a cell as a distribution of molecules and of the cell cycle and adaptation in general as changing the this distribution through time, thermal adaptation also by definition requires a change in the distribution of molecules in the cell. The Thermal Code clearly provides the mechanism for this. Although the thermal dependence for activity of proteins is obvious, a small set of finely tuned regulatory proteins can not possibly be responsible for the general adaptation to the thermal environment problem given the nature and complexity of the cell cycle.

    This is a very important point respecting validation of the Thermal Code construct but is not immediately obvious. Thermal adaptation like timing problems and other important and fundamental problems in biology have been poorly to not at all understood phenomenon, so using an explanation of these phenomenon through the Thermal Code construct as a validation of the construct first requires getting into touch with the meaning and significance of these exceedingly fundamental problems.

50. Discuss timing problems and their relation to the Thermal Code, nucleic acid mainframe model.

    Time in biology is measured in terms of concentration. The cell can be viewed as a battery, different pools of small molecules being built up through time so measuring the concentration through time of various small molecules and their rate of change by some molecular mechanism of sufficient resolution capability and which can effect “decisions’ in response to this information of concentration cam allow a solution of the general “timing” problem in biology. The Thermal Code/nucleic acid mainframe provides an explanation for the timing problem in a clear manner. Providing both a definition of and mechanism for the timing problem in general is an absolute requirement for a holistic theory of the cell. This is an exceedingly strong motivation for and validation of the Thermal Code/nucleic acid mainframe model. It would be difficult to over emphasize this. The span of time scales over

which regulation is obtainable extends to the cell cycle through the Thermal Code/nucleic acid mainframe model in a simple way.

51. Discuss the spatial configuration of RNA polymerase, its behavior relative to the DNA ds, and coupling of the collision parameters to the Thermal Code.

    The RNA polymerase is homogeneously distributed throughout the nucleoid. It is there because it interacts with the DNA ds, among the host of factors which lead to phase partitioning. The σ subunit is not necessarily restricted just to the nucleoid and there is evidence it is found both in the cytoplasm and nucleoid. Some regulatory function respecting the σ monomer of the holoenzyme, etc, can be obviously assumed and defined by the Thermal Code/mainframe model.

    A reasonable interaxial distance between ds of 50 – 100 Å and the size of the polymerase implies many contacts and restricted diffusion for the polymerase in the DNA matrix of the nucleoid.

    Since the DNA in the nucleoid is almost entirely supercoiled, polymerase can be envisioned to be surrounded by bundles of supercoils.

    So the polymerase random walk in the condensed nucleoid phase we can imagine that rotational motion of the polymerase dominates translational search over the ds. The supercoils are in dynamic equilibrium with all the implied DNA translational and rotational motion, so a large amount of the "search work" is contributed by the DNA dynamics, not a one dimensional diffusion of RNA polymerase on the helix.

    So the equilibrium configuration of the RNA polymerase/DNA in the nucleoid provides a continuity of contacts through thermal fluctuations without a severe departure from the random looking equilibrium configuration, i.e. the minimum entropy production for search.

    For the Thermal Code to operate requires a coupling between collisions or contacts of the RNA polymerase with the DNA ds and the formation rate of signals (destabilized zones) on the ds via the Thermal Code. This requirement is similar to that of no monomer gradients in the nucleoid, information processing is impossible without these requirements. The proper collision rate follows in a simple way from the concentration of RNA polymerase and DNA and the phase definition. Since it is the equilibrium looking configuration which provides the background for a clean information processing system through the Thermal Code/mainframe construct, its maintenance is automatic or continuously restored from various non-equilibrium configurations by the global equilibrium and equilibrium seeking/perturbation cycles already discussed.

52. Why so damned much discussion of equilibrium, equilibrium appearing configurations, equilibrium seeking configurations, etc?

Virtually all of thermodynamic has been constructed based on the simple observation of a direction towards some stable, non-evolving, equilibrium. It is clearly recognized that all systems of matter evolve towards equilibrium. The role of this construct in physical science can not be overstated.

The most central problems in understanding "Life" revolve around their non-equilibrium behavior and configuration, the neg-entropy flow system jazz from biophysics.

The Thermal Code/mainframe model effects an incredibly powerful information processing system which is all downhill chemistry. This molecular system motivates information gathering and processing as downhill chemistry. We can start with a configuration of the cell and by imposing small molecule non-equilibrium configurations (through mass-energy exchange with the environment, converting to monomers in the cytoplasm, which will be discussed in the same framework to extend the information gathering all the way from DNA to the environment as just downhill chemistry later on) obtain a powerful regulatory architecture leading to a distribution of ss through time from the nucleoid. This is the central riddle of Life and its thermodynamic solution. The level of existing quantification and description for molecular systems currently around do not approach the potential of this construct. Suffice it to say at this point that because of the Thermal Code it will be possible to completely quantify the chemical micro-environment and a different computational and conceptual approach will be needed. A lot more on this later. The Thermal Code/mainframe will allow mapping from microscopic to macroscopic with 0 information loss as a physical chemical experimental system it is astonishingly powerful. There are nice philosophy reasons why this must be the case with DNA.

53. What is required of the monomer evolution in the nucleoid?

    It must contain the information of the cell condition.

54. How does this obtain in the cell?

    Various processes and enzymes use monomers as substrates or allosteric effectors. The most important division of labor for monomers is the specific monomer used as activators or carriers for membrane, cell wall, protein synthesis system and general synthetic pathways.

    Furthermore key proteins in the small molecule net are highly tuned or have rate behavior which is a function of various monomer concentrations. The Thermal Code provides the basis for coupling these diverse events in a powerful way.

55. How are we to view proteins in this regulatory scheme?

For every protein it has a reaction profile over the concentration space of substrates and effectors.

Given this description of the proteins behavior the Thermal Code/nucleic acid mainframe role is then to determine the # of these proteins in the cell, of course what being important is the relative 3's of proteins in the cell, precisely what the DNA sequence programs.

56. What about spatial configuration in the cell?

    The phase description and flow system or dynamic configuration provide the different chemical environments for localization, partitioning, or random walk path for the proteins.

57. In viewing the cell then, proteins can be thought of simply as tools, effecting only "regulation" in a small local environment, completely subordinated to the nucleic acid chemistry respecting regulation.

    Yes. They are just simple catalysts. rather than viewing them as effecting regulation, they can be viewed as being regulated by their environment. There are obviously a range of interactions with the environment from decoupled and insensitive to composition, to highly sensitive to a number of different aspects of the environment.

58. How can this demotion of proteins fit in with evolution?

    Very nicely for obvious reasons. Proteins are tools "required" by the nucleic acid system.

59. The word environment or composition has been used frequently throughout this discussion, exactly what is meant?

    The environment around any macromolecule or in any part of the system <u>is simply a list of composition</u>.

    1) H_2O
    2) H^+
    3) Mg^{++}
    4) ... all small molecule, ion, etc, species.

    The <u>temperature</u> must also be specified.

    And this list of components is inside some volume, confined to <u>a region of space</u>.

    This is a complete description. Viewing the chemical system in this way will be essential later to building up a computationally tractable description of the chemical

phenomenon in the cell at various hierarchical levels, includes all the relevant information and is highly amenable to the crunching or quantification process.

Different sized systems can span from around a given macromolecule, a phase, to the whole cell. Many different ones will be chosen for various reasons. Mass-energy exchange over the boundaries of each system viewed (the exchange over the nucleic acid mainframe boundary will be of obvious importance) will be included.

60. Describe the tree. this just the first, crudest, conceptual pass. When finished the tree and its change in time will be totally quantitative biology → all chemistry.

    All chemical species which are in the cell will be included.

    The tree is built by putting arrows between all reactants and products.

    Associated with each chemical species will be the instantaneous concentration of that species.

    So any chemical reaction will have associated with it a change in the # of the relevant reactants and products. H_2O, H^+ generation or use are of course included.

    Associated with each arrow will be a list of the proteins and other chemical species which catalyze or are catalytically involved. This includes pH → just H^+ #, ionic → a # → all just #'s, effector molecules, etc. The concentration of all these species are part of the tree.

    → Defer question until simpler view of list is complete.

61. Why will it be possible to obtain a totally quantitative biology at the level of the prokaryotic cell?

    Because of the Thermal Code/nucleic acid mainframe information and transport system. The basic point is regulation via the Thermal Code is systematic.

62. How much of the RNA made is initiated through the Thermal code?

    ~100% by definition of the RNA polymerase/DNA interaction mechanism (with some possible exceptions).

63. What about regulatory proteins?

    The regulatory architecture defines their action to be exerted at two levels, the # ss made cycle and the # proteins translated cycle.

    The range of interactions at the transcription level spans from irreversible blockage of the DNA destabilization Thermal Code pattern to irreversible facilitation of

destabilization – wide open. Intermediate reversible interactions around and with the promotor can compete for both structural and sequence specific signals.

The general notion is that when the regulatory protein is not around the DNA segment of interest the Thermal Code is operating. When it is around, competition and modification of the RNA polymerase interactions can be effected.

The role of the degree of reversibility of the regulatory protein/DNA interaction is obvious; with a specification of the protein/DNA segment collision rate and residence time of binding, if we already know the RNA polymerase/DNA collision rate and Thermal Code interaction parameters, the total transcription regulation can be quantified.

The special problems which could arise can be listed and quantified. A protein which requires transcription and turns on operation of that DNA segment by leaving and allowing the Thermal Code to operate will be easy to handle. Proteins that go on and off and compete with RNA polymerase for Thermal Code generated signals and sequence specific interactions can also be handled in a straightforward fashion, in a similar way in which the RNA polymerase/DNA interactions can be understood, we just add competing species. However proteins that bind to the DNA and cause a destabilization event adjacent to them as a signal for the RNA polymerase/DNA interaction, and/or interact directly with the RNA polymerase to alter its collision parameter or mode of attack with the DNA could be difficult to model in a systematic way, each requiring its own set of empirically derived parameters in order to obtain a quantitative model. The problem here is complete decoupling from the Thermal Code signal generation/RNA interaction cycle. This is not as bad as it may first seem since the regulation of these segments of DNA must still be coupled to the cell cycle, the regulatory proteins themselves must come form somewhere in a quantitative cell cycle dependent fashion, the transcription of these segments must always have a relation to the distribution of ss being made, so it is just a problem of having to dig out some specific empirical parameters which then will allow integration into the systematic regulatory architecture.

64. How much delineation of the protein regulatory mechanism at the transcriptional level is obtainable through knowledge of the structural products ultimately derived from those segments of DNA?

    Quite a bit. Many general rules emerge.

    1) Proteins are only used in regulation when the Thermal Code cannot do the regulation. This can arise because the information cannot be encoded through the monomer concentration language or the magnitude of regulation cannot be obtained through the Thermal Code channel.
    2) Small molecules provided in the environment which reduce the necessary transcription from the normal Thermal Code regulated level would require a

3) Any entity which is decoupled from the cell cycle may have a regulatory protein involved at the level of transcription. Some small molecules it may simply be desirable to keep at a certain level. However Thermal Code promotors can also be decoupled from the cell cycle and still coupled to global production rate of ss.
4) Different substrate utilization from the environment may require protein involvement.
5) In general, when the information of relevance cannot get through the Thermal Code channel a protein becomes implicated.

65. What is the general nature of regulatory proteins at the level of transcription then?

If the Thermal Code can't do the job a protein becomes implicated. These can override, modify or decouple Thermal Code regulation. But in all cases they have a relation to the main regulatory scheme. Empirical parameters are needed to elucidate and quantify the relation.

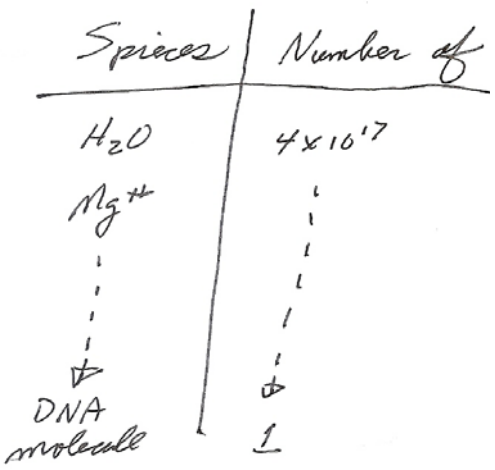
66. Give the simplest first view in developing the quantification scheme for the cell.

We simply have lists of instantaneous composition of the cell. We have a list for different times through the cell cycle. To resolve the macromolecular synthesis will require time separations related to transcription initiation and polymerization rates, translation initiation and polymerization rates; on the order of one second.

The list is complete. each species inside a boundary over the cell is included. Connectivity determines the different species. H_2O, H^+, are included. Everything that's in there.

We won't be concerned yet with the practicality of building these data structures. First we will build up to certain structures and levels of description before we can force quantification from known obtainables.

The list looks like:



67. What information is in a sequence of these lists taken at points through the cell cycle?

    With the temperature and the volume defined by the boundary these lists have all the information in the system. A given spatial organization is assumed at the start or for the first list. This organizational information will represent additional information not included in the list if the system cannot self-organize (kinetic, mechanistic barriers). from the isotropic configuration. If the system could self-organize from the isotropic configuration of the species then temperature, volume and the list would have all the information in the system. There are very interesting subtle points here which are being ignored (involving entropy of and mechanism for organization).

68. What information is available by watching the change in concentration in a sequence of these lists?

    The Δ's give the mass-energy flow through the system. Running a cell is equivalent to changing the concentration of species in the list. Everything is in there, all regulatory architecture, enzyme kinetics, diffusivities, etc, the problem is to get out the relationships.

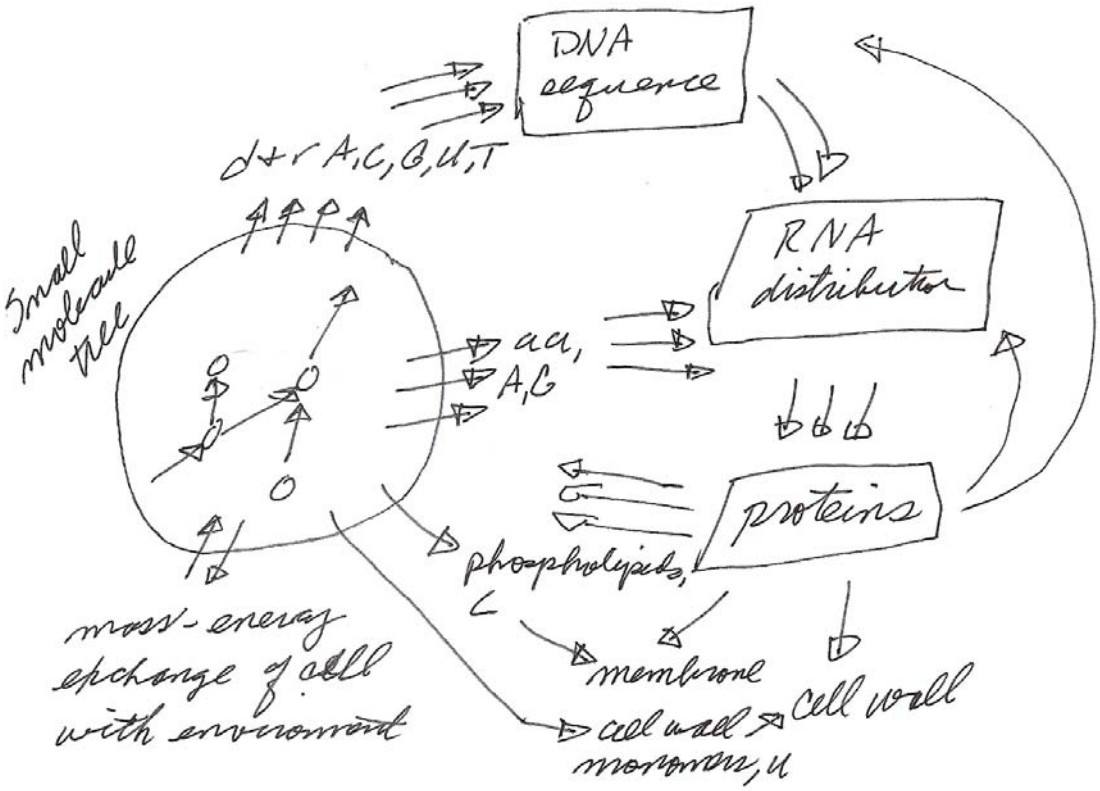
69. Extend the lists to general trees.

    The tree is similar to the list in that it has all the species concentrations of. But they will be averaged in the tree such that an arrow connects all reactants and products.

    Associated with each arrow will be a list of all catalytically associated factors. This includes the protein catalysts, H_2O, H^+, small molecules, etc. The list of associated factors is like the other lists, giving species and concentration of. These arrow associated lists are no more than a special collection of the data in the cell list/time.

The reactants, products and arrow associated list between will define a point in the reaction profile of the protein in concentration space in vivo which catalyzes the reaction. The Δ numbers between successive trees yield a simple in vivo reaction rate. Transport problems will have to be included before the simple reaction rate can be decomposed kinetically to diffusion, effector binding, pure chemical reaction rate at the catalytic site, etc. The spatial/temporal organization in the cell will be necessary to get to this level of description of each individual reaction.

70. Sketch a crude tree of the cell.



The organization of the cell is such that the transport implied by the arrows is physically minimized; the minimum transport function. The presumed organization has already been outlined.

71. Decompose the tree into the phases of the cell.

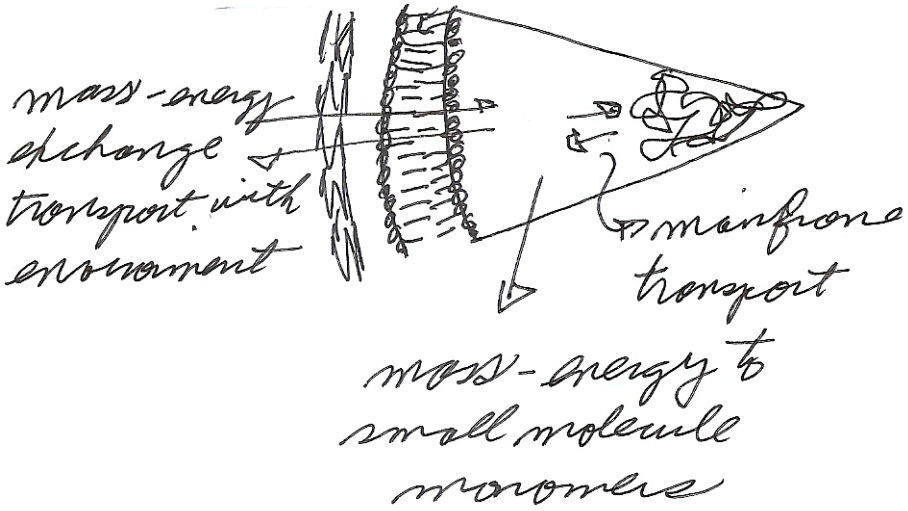
Each phase is defined spatially by geometry and symmetry and has a composition which evolves over the binary fission cycle and undergoes continuous mass-energy exchange over the boundaries with contiguous phases.

The diffusion between what goes in and what comes out of each phase yields the composition time evolution of that phase. The fission cycle implies the boundary.

The spatial definition and time evolution for each phase will imply the transport parameters, partial molar volumes, etc, for all the species in each phase. This will

include all ionic and other gradients etc that occur. relating the different parameters will lead to complete understanding of the dynamic behavior.

So for each phase there will be a list of composition, the changes in lists through time implying the composition time evolution and transport through.



Five phases with composition lists through time:

1) environment
2) cell wall
3) membrane
4) cytoplasm
5) nucleoid

The stoichiometric balance required is obvious. The nucleic acid mainframe/Thermal Code is the information processing system which effects this. The knowledge of phase composition and their time evolution and transport must be the information processing and coding scheme by definition of a cells existence with the DNA molecule as memory.

72. Unfortunately we do not have the information needed to fill in these trees. How can we get it?

There is only one reason and one way to force the quantification. We will have to have the Thermal Code interactions in a quantitative form and have the DNA sequence so we can simulate the transport over the nucleoid interface. The regulatory proteins will have to be included and it is believed these can be handled by obtaining the specific empirical parameters for their interface with the DNA. The important point however is that the Thermal Code interactions will constitute the majority of the RNA synthesis. This has already been discussed.

The general scheme than is to start out at the root of the regulatory structure, Thermal Code interactions, which predict and should allow complete quantification and work out, eventually to the mass-energy exchange over the cell boundary.

Only the mainframe requires a perfect or complete simulation. Downstream events will be treated statistically. Obviously by the time we are to something like small molecule production or H^+ pumping, counting each event will be unnecessary and the large numbers can be handled statistically and by symmetry based sectioning.

The 2°/3° RNA determining (and protein feedback on 1°) can probably not be approached analytically. 2-d quantitative O'Farrel's using 1° (from DNA sequence) to calculate molecular weight and electrolytic behavior on synchronous cultures coupled with mainframe simulation will allow messenger 2°/3°to be correlated to protein translation number. Statistical analysis, etc., will probably greatly simplify the seemingly unapproachable 2°/3° ss problem. The data base forces at least an statistical solution to the problem.

Cell wall and cell membrane should be easy to quantify monomers for due to homogeneity.

With the nucleic acid, proteins, cell membrane and cell wall data through time, the monomer requirements through the small molecule tree are quantified.

Changes in all pool size data through time will need to be estimated to get to the actual concentrations from known throughput data implied by the macromolecular synthesis.

The small molecule concentration data through time imply the mass/energy exchange over the cell boundary by simple stoichiometry.

So the general sequence to be able to build the tree is

1) break out Thermal Code as *f*(monomers, sequence)
2) add in regulatory proteins (1 & 2 RNA distribution/t)
3) need dna sequence
4) O'Farrel 2-d quantify by knowing 1°
5) get translation #/message (RNA 2°/3°)
6) add in quantification of cell wall and membrane
7) then can force whole thing.