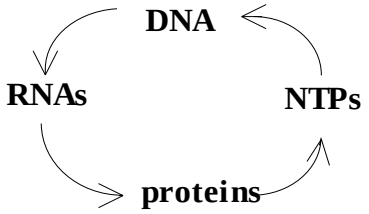
What Life Is[1]

David Martin Degner
East Anchorage, Alaska

Abstract: A static and dynamic physical model is presented for the Gram (+) prokaryotic cell, the hydrogen of biology. Two currently unknown models that I have named the **Nucleic Acid Mainframe** and the **Thermal Code** are presented here. They are the most primitive and most important models in the origin of life chronology. Combined with the "RNA world" origin of life scenarios, these models can be seen to be the first two "missing links" in that origin of life evolutionary path. Together with the Genetic Code and the largely known complex network of biochemical reactions in biology we have the answer to the seminal question—What is Life? It is discovered that cells are chemical computers and the fundamental Turing machine architecture of biology is described here. The condensed DNA of prokaryotic cells forms a separate phase, the nucleoid phase, surrounded by the cytoplasm phase. The RNA polymerases (RNAPs) are distributed randomly throughout the nucleoid. rNTP monomers (ATP, GTP, UTP, CTP) flow into the nucleoid from the cytoplasm and single strand RNAs originating from random locations throughout the nucleoid where transcription is taking place flow out of the nucleoid into the cytoplasm. I call this vectorial flow process between the nucleoid and the cytoplasm the Nucleic Acid Mainframe. While inside the nucleoid the newly synthesized RNAs are largely restricted from folding to form secondary and tertiary structure due to restriction by the DNA matrix and do a random walk to the surface of the nucleoid phase. At the surface of the nucleoid single strand RNAs can fold to form secondary and tertiary structure and interact with proteins and ribosomes in the cytoplasm. I call this dynamic surface region where RNAs are folding and interacting with proteins and ribosomes the **RNA Processing Zone**. The Thermal Code model postulates that the "state" of the cell is encoded in the list of rNTP monomer concentrations that evolve over the cell cycle. The Thermal Code model postulates that the global level of transcription is determined by the toroidal supercoil density that is determined by the ATP concentration via ATP utilizing topoisomerases. The Thermal Code model postulates that the differential rate of transcription at an instant in time, the pattern for transcription at an instant in time of the different genes, at constant toroidal supercoil density, for many but not all genes, is determined by the four rNTP concentrations and specific DNA sequences the monomers interact with. The Thermal Code is coupled to the binary fission growth cycle. Then the linear **Central Dogma** becomes a closed circle or loop.



Introduction

A precise physical description of a cell would be to specify the structure and position of every atom, ion, and molecule and their evolution over time. The position means position in space and changes over time. In biology, classical space and classical time are fully able to define position and motion. The structure is the quantum mechanical structure. Atoms, ions, and molecules exhibit two phenomena of primary interest—diffusion and reaction. Atoms, ions, and molecules diffuse slowly to a specific configuration, such as a small molecule to the active site of a protein, and then a reaction occurs fast, virtually instantaneously. Diffusion and reaction are the physical chemistry game of cells.

[1] Sometimes we can't see the forest for the trees. This paper describes the forest.

There is another game of cells, the only game, that I call the concentration game, that is an information/counting game. What a cell is at a fundamental level is a collection of atoms and molecules, from the smallest, the proton, to the largest, the DNA, and the key information is how many of each of those entities are in a cell, their concentrations, that evolve over the cell cycle. That is regulated by replication, transcription and translation, the central dogma of DNA to RNA to protein.

A cell is a complex chemical kinetics net. The rNTP monomers are both information and energy carriers. Reaction rates are a function of the concentrations of the reactants. Reaction rates are the mechanism by which rNTPs exert kinetic control of the cell chemical kinetic net.

The most important level of organization of the cell is the phase organization. The definition of phase I use is a region of space that has constant composition net of thermal fluctuations. The information of a phase is the size and shape of the phase in space and the list of concentrations of the component atomic and molecular species. Transport in a phase, in the absence of a macroscopic electric or magnetic field defined on the phase, is strictly limited to random walk diffusion. Transport between phases can be vectorial, non-random, into one phase and out of the other phase. The transport between phases in cells is vectorial.

The gram (+) prokaryotic cell, the hydrogen of cells, is composed of 4 phases:

1. Nucleoid.
2. Cytoplasm.
3. Membrane.
4. Cell wall.

Cell phases have different inside and outside surfaces. The nucleoid has no inside surface, it is at the center, and has a nucleoid/cytoplasm interface. The cytoplasm has an inner interface with the nucleoid and an outer interface with the inner membrane surface that are different. The membrane has inner and outer surfaces that are different. The cell wall has inner and outer surfaces that are different, there is organization over and above the strictly phase definition. This enables and reflects vectorial flow of ions and molecules between phases.

The Nucleic Acid Mainframe

A fully quantitative description of the nucleoid cannot yet be made. The size and structure of the nucleoid, the DNA toroidal supercoil density and dynamics, the DNA helix/denaturation bubble, kink, hinge, random coil equilibrium, the number of RNAPs, the nucleoid associated proteins, the role of protein/DNA binding, protein/DNA bending, the topoisomerases that put in and take out supercoils, are not known with quantitative precision. It is a complex system. In this system resides the Turing machine of life. The DNA, the finite one-dimensional tape of the Turing machine, is searched to determine a transcription profile. The transcription profile is simply the list of RNAs being synthesized at an instant in time. The RNAPs are the biological Turing machine heads—cells are multi-head Turing machines. What is the spatial/temporal description for this system?

What is the composition of the nucleoid? A typical bacterial nucleoid contains $10^{6\text{-}7}$ base pairs, and $10^{3\text{-}4}$ genes. Ions, small molecules and proteins partition between the cytoplasm and the nucleoid. The replication and transcription machinery are in the nucleoid. Not in the nucleoid are all the translation and metabolism machinery found in the cytoplasm. Leaving out the replication machinery, transcription regulatory proteins and other nucleoid associated proteins the composition of the nucleoid is a relatively simple four-part system composed of DNA, RNAPs, rNTP monomers, and freshly synthesized single strand RNAs. The RNAP holoenzyme, $\alpha_2\beta\beta'\sigma^{70}$, is big, 100 × 100 × 160 Å, and massive, 450KDa, the same mass as ~ 680 base pairs of DNA. There are $10^{3\text{-}4}$ RNAPs per nucleoid. A significant fraction of the volume and

mass of the nucleoid is RNAP. In an E. coli with 3000 polymerases and 4.6×10^6 base pairs of DNA, one genome, the mass of the RNAPs would be 44% of the mass of the DNA. Exponentially growing E. coli have 4.4 genomes in the nucleoid, enabling the shortest binary fission time to be less than the replication time for one genome. Because the polymerase is compact and the DNA is extended the polymerase is a smaller volume fraction of the nucleoid than mass fraction.

The DNA in the nucleoid is often negatively supercoiled but can be positively supercoiled in thermophilic bacteria. The forces that give rise to the nucleoid phase are mechanical forces and electrostatic forces. Mechanical forces exist in molecular connectivities, like the covalently bound DNA backbone, where mechanical forces can be transmitted up and down the DNA molecule. Mechanical forces also exist between molecular connectivities and mechanical forces are exchanged between DNA and RNAPs in the nucleoid. Electrostatic forces are electric field forces that arise between ions and atomic dipoles, plus and minus trying to get as close together as possible, plus/plus, and minus/minus repelling each other. Hydrogen bonds are an example of electrostatic forces. The DNA in the nucleoid of bacterial cells is highly condensed due to three factors:

1. Toroidal supercoils.
2. Denaturation bubbles, kinks, hinges, random coil sequences.
3. Interactions with RNAPs.

There are attractive electrostatic forces between DNA and the RNAPs, that's why the RNAPs are in the nucleoid surrounded by DNA. A significant electrostatic force is that between the negatively charged oxygens on the phosphodiester backbone and Mg^{++}. rNTP and dNTP monomers carry significant charge and are closely associated with Mg^{++}. Proteins with positive charge that can interact with the negatively charged phosphodiester backbone partition into the nucleoid.

Supercoils come in two topologically equivalent conformations: interwound supercoils and toroidal supercoils. A negative supercoil corresponds to a right hand interwound supercoil and a left hand toroidal supercoil. I believe only toroidal supercoils occur in vivo because interwound supercoils would be anathema for the transcription model I am proposing and would be anathema for replication. Interwound supercoils do occur in vitro and there are electron micrographs of interwound DNA. The specific linking difference (the linking number minus the relaxed level of twist divided by the relaxed level of twist) in E. coli is −.06. If three-quarters of the linking difference goes into negative supercoils and one-quarter into negative twist, and we assume 10.5 base pairs per twist in relaxed DNA, then in an E. coli with 4.6×10^6 base pairs of DNA there would be 19,714 left hand toroidal supercoils and the supercoil density would be −.0043 supercoils per base pair. If each supercoil is composed of 125 base pairs 54% of the DNA would be supercoiled helix and 46% non-supercoiled helix. In general, in typically negatively supercoiled bacteria there are on the order of $10^{4\text{-}5}$ supercoils in a nucleoid with $10^{6\text{-}7}$ base pairs DNA.

Where does transcription take place? The RNAPs are distributed randomly throughout the nucleoid. Transcription takes place inside the nucleoid at random sites throughout the nucleoid. The idea that transcription takes place on the surface of the nucleoid at the interface of the nucleoid and cytoplasm is not physically plausible. Mechanical forces are exchanged between the DNA and the RNAPs determining their two partner "dance". The RNAPs open a hole in the nucleoid DNA matrix and are confined to a fixed position in the nucleoid by contacts on all sides with the DNA matrix. The "dance" is the partitioning of the rigid helix, rod-like, toroidal supercoiled sequences to between the RNAPs and the partitioning of the bendy denaturation bubble, kink, hinge, random coil sequences to the RNAPs. During transcription elongation, the RNAP must remain fixed in place and the DNA translates through the RNAP/DNA/RNA ternary complex. The RNAP is very massive so it does not translate, rather the DNA that has low linear mass density must account for the relative motion between RNAP and the DNA. Since the DNA is a helix it must also twist though the fixed RNAP/DNA/RNA ternary complex. In a bacterial cell with negatively

supercoiled DNA collapse of a left hand toroidal supercoil in front of the RNAP/DNA/RNA ternary complex coupled to formation of a left hand toroidal supercoil behind the complex allows the DNA to twist through the fixed ternary complex. The toroidal supercoil density remains constant net of thermal fluctuations but the toroidal supercoils are moving around on the DNA.

All the DNA in the nucleoid is equivalent and this is required for kinetic regulation to be possible. The toroidal supercoils and the bubbles, kinks, hinges, random coil sequences are seeking equilibrium, forces flowing freely along the DNA, the DNA twisting, toroidal supercoils migrating around with the energetics determining the local structure. This complex state is one of seeking torsion equilibrium over the entire DNA molecule, with the processes of transcription and replication perturbations from equilibrium, they are non-equilibrium processes. In the stationary growth phase, where there is no transcription and replication taking place, the entire nucleoid is in equilibrium.

The RNA single strand leaving the RNAP/DNA/RNA ternary complex does a random walk to the surface of the nucleoid. While on this random walk the RNAs are largely prevented from folding by the nucleoid DNA matrix. When the single strand RNA gets to the surface of the nucleoid it can fold to form secondary and higher structures. rNTP monomers do a random walk in the nucleoid to the RNAP/DNA/RNA ternary complexes where synthesis is taking place. Overall two vectorial flows occur between nucleoid and cytoplasm associated with transcription.

1. rNTP monomers diffuse into the nucleoid from the cytoplasm.
2. Single strand RNA diffuses out of the nucleoid into the cytoplasm.

I call this flow system the Nucleic Acid Mainframe. It is what is called in non-equilibrium thermodynamics a dissipative structure. The high energy of the rNTP monomer pools is the source of energy that drives this flow system, that energy being converted to lower energy RNA polymers and phosphates. At the interface between the nucleoid and the cytoplasm is where the RNA single strands fold to form secondary and higher structures and can begin to interact with proteins and ribosomes in the cytoplasm. I call this dynamic surface region that arises when transcription is taking place the RNA Processing Zone.

There are important molecular evolution and origin of life implications of the Nucleic Acid Mainframe model. Starting with DNA and counterions the nucleoid phase can self-assemble and rNTP and dNTP monomers spontaneously partition into the nucleoid. Polymerization of RNA using ribozymes as crude RNAPs, part of the RNA world early evolution scenario, can lead to primitive transcription. Then a transcription system precedes a translation system and proteins. The associated transcription initiation model, the Thermal Code, has implications to a primitive transcription system that synthesizes RNAs. Of course, given the complexity of a cell, it is always a bit difficult to understand how such a system could spontaneously arise.

The Nucleic Acid Mainframe is not available in current in vitro systems. The nucleoid is a highly concentrated and exquisitely balanced system only found in vivo. The concentration of biomolecules is very much higher in vivo than in vitro. It may be difficult or impossible to assemble and run a nucleoid in vitro, then again, it may be possible starting out with my model.

Introduction to Transcription Regulation

The general transcription regulation problem in bacteria is how to turn on and off 10^{3-4} genes in a precise quantitative way. There is both the problem of setting the global transcription rate by a nucleoid and the differential transcription pattern. The time period of regulation spans at least 3−4 orders of magnitude from initiation every ~ 1 second to initiation every 10^{3-4} seconds.

How the DNA is searched to arrive at the transcription pattern determines how powerful of a Turing machine the cell is. The key reaction is transcription initiation. The transcription pattern is determined by differential transcription initiation for different genes. Therefore, transcription initiation determines how powerful a computer the cell is. How is the DNA searched to arrive at a transcription pattern?

Replication, transcription, and recombination require unwinding of the DNA double helix. In all three processes one or both strands must have the H-bonding faces of the bases rotated out from the helix configuration to facing into the surrounding solution. This is required for the DNA single strand to serve as a template where it can H-bond with incoming bases either as monomers in transcription and replication or with another single strand DNA in recombination. Where on the genome is the DNA helix unwound and why? Is it only unwound in combination with proteins?

RNA polymerase/DNA Collisions

Consider collisions between RNAP holoenzyme and DNA. Of course, collisions between RNAP and DNA are complex and not simple like a collision of billiard balls. There are two possible outcomes to a collision—initiation or rejection of initiation. At constant supercoil density there are five possibilities that can determine collision outcome between RNA polymerase and DNA:

1. Sequence determines the collision outcome. This would involve specific sequence recognition in the major and/or minor groove.
2. A structural feature of DNA such as a disrupted helix determines the collision outcome and this structural feature is determined by sequence alone.
3. A structural feature that is a product of the interaction of DNA with transcription regulatory protein determines the collision outcome—this is obviously very important, like the lac operon.
4. A structural feature that is the product of interaction of DNA with rNTP monomers determines the collision outcome—what I am proposing in this paper.
5. Some combination of the above—the actual case.

The Constant Code

For scenarios one and two above there is no information processing. The DNA sequence alone determines the collision outcome at constant supercoil density. The RNA coding regions of genes give a rejection of initiation. The promoters have a probability of initiation associated with the RNAP holoenzyme/DNA collision. This requires a code. I will call this code, based on sequence only at constant supercoil density, the Constant Code, because in the absence of transcription regulatory proteins there is no way to turn genes on and off—there is no way to change the differential pattern of expression. But the global transcription rate can be determined through the supercoil density. The Constant Code determines the in vitro strength of promoters, the open complex formation pattern.

The Thermal Code

I have two mechanisms postulated for the Thermal Code model. Both postulate that the state of the cell is encoded in the list of rNTP monomer concentrations.

1. The first Thermal Code model mechanism further postulates that in the nucleoid of prokaryotic cells interaction of rNTP monomers with the denaturation bubbles, kinks, hinges, random coil sequences in the DNA lead to dynamic hybrid helices composed of monomers interacting with the DNA single strand of the denaturation bubbles, kinks, hinges, random coil sequences. These hybrid helices are hypothesized to be stochastically defined, meaning they fluctuate over time, only having

a smeared-out existence, and are hypothesized to be centered on the Pribnow boxes. In this Thermal Code scenario, the diffusion to the RNAPs of these structures are hypothesized to be the critical kinetic intermediates for transcription initiation in the absence of regulatory proteins.

2. The second Thermal Code model mechanism postulates that the key regulatory chemical event is the kinetics of the first several rNTP monomers being polymerized into RNA on the DNA/RNAP binary complex. This is guaranteed to have a reaction rate that is a function of the rNTP concentrations but not guaranteed to be the key regulatory events.

Both mechanisms can exist or the first mechanism can exist without the hybrid helices but with diffusion of the bubbles, kinks, hinges, random coils to the RNAPs as a key kinetic regulatory element. I think both mechanisms for the Thermal Code exist as described and work in unison.

Genetic Code Degeneracy and Codon Usage

The most important aspect of the genetic code is degeneracy—multiple codons per amino acid. Codon usage statistics reveal that most codons are used at significant levels. Wobble allows a smaller number of classes of tRNAs than codons to be used for translation. The reason degeneracy is the most important feature of the genetic code is that it allows a transcription initiation/rejection of initiation code to co-exist with the genetic code amino acid specifying function of DNA. Protein coding regions must be coded differently than promoter sequences and that is precisely what degeneracy allows. In addition to protein coding sequences the leader sequences, tRNA sequences, and rRNA sequences must be coded not to be promoters—to give rejection of initiation on collision with polymerase. The words that determine initiation or rejection of initiation are sequences, N-mers, of unknown length N. N is probably more than 5 bases and less than 15 bases long and probably is variable in length along the DNA sequence. The number of words grows fast as 4^N. RNA coding sequences of the DNA are the subset of the 4^N large set of words that give rejection of initiation on collision with polymerase. Promoters are the subset of the 4^N large set of words that give initiation on collision with polymerase. Degeneracy and codon usage are strong evidence for a transcription initiation/rejection of initiation code in DNA primary sequences.

Setting the Global Transcription rate

The global rate of transcription is set by two factors:

1. First the energetics of how difficult it is for RNAP to separate the DNA strands and get “into” the helix.
2. And second that energetics also determines the rate of denaturation bubbles, short sequences of denatured DNA, kinks, hinges, random coil sequences.

Consider the relative motion of RNAP and DNA in the nucleoid. The problem of searching the DNA to determine global and differential expression is either done by the RNAP diffusing along the DNA or the DNA diffusing to the polymerase. Because of the large mass, 450KDa, of the RNAP and the low mass of DNA, 660 Da/base pair, the RNAP does not do a topologically complex one-dimensional diffusion along the DNA helix. In the nucleoid the polymerase is relatively fixed in space by numerous contacts with supercoiled condensed DNA. The DNA is twisting and supercoiling, transmitting energy and forces along the backbone, and supercoils are migrating around. The DNA is in motion—the RNAP is stationary. As in elongation where the ternary complex is fixed in the space of the nucleoid and the DNA translates through the ternary complex, for initiation the RNAP is fixed in the space of the nucleoid and the DNA does the diffusion to the polymerase. Collisions between a given sequence and the RNAP are not random over time. In the nucleoid the denaturation bubbles, kinks, hinges, random coil sequences assort to the RNAPs and the rigid helix, rod-like, toroidal supercoiled sequences assort to between the RNAPs. This partitioning process

in part determines the transcription rate, both global when taken in aggregate and differentially when looking at individual promoters. Three factors determine the energetics:

1. Sequence: AT rich sequences denature first, like the TATAAT Pribnow box.
2. Temperature.
3. Toroidal supercoil density: The more negative toroidal supercoiling, the easier it is for the RNAPs to get into the helix and the higher the level of denaturation bubbles, kinks, hinges, random coil sequences.

The sequence and temperature are fixed so the only regulatable factor is the toroidal supercoil density. That level is determined by the ATP utilizing topoisomerases that put in either negative or positive supercoils. In most bacteria the DNA has negative supercoiling but in thermophilic bacteria that are GC rich the DNA can be positively supercoiled. The rate of reaction of the ATP utilizing topoisomerases is a function of the ATP concentration, so the density of toroidal supercoiling is a function of the ATP concentration. The DNA is "searched" for promoters by the partitioning process, where the DNA denaturation bubbles, kinks, hinges, random coil sequences diffuse to the RNAPs. The time it takes for that diffusion process is a kinetic regulatory element of transcription. The Pribnow box sequences and adjacent sequences are the place where those helix disruptions occur. To put in toroidal supercoils, either negative or positive, requires work and an energy source that is provided by the ATP used by the topoisomerases. Letting toroidal supercoils out and reverting to the relaxed state does not require energy, rather releases stored energy. ATP is the primary energy and information molecule in the complex kinetic net of enzymes in the cytoplasm so setting the global rate of transcription by the ATP concentration makes sense and is an essential part of the Thermal Code model.

The Hybrid Helix Thermal Code Mechanism

One mechanism for initiation I am proposing postulates the existence of a DNA structural feature, namely, hybrid helices of rNTP monomers and single strand DNA. I believe that in the nucleoid there are denaturation bubbles, kinks, hinges, random coil sequences migrating around as the DNA twists and supercoils. rNTP monomers can interact with these bubbles, kinks, hinges, random coil sequences forming dynamic hybrid helices hypothesized to be of length 3−12 bases on one or both strands of the bubble, kink, hinge, random coil sequence. The rNTP monomers can make both stacking interactions and conventional H-bonding in these hybrid helices leading to favorable energetics. I think these structures are centered on the −10 Pribnow box and the adjacent sequences. These structures are dynamic in that they are bendy regions of DNA and the monomers are making and breaking H-bonds with the single strands of DNA. At any instant in time the number of pairings is stochastic—they have an average structure of so many base pairings. It is this average structure that I call a hybrid helix. I believe the rate limiting step for initiation in the nucleoid is the diffusion of these structures to the RNAPs. The longer the length of the hybrid helices the faster the diffusion to the RNAPs. A hybrid helix is a bendy stretch of DNA and the rest of the DNA helix, albeit toroidal supercoiled, is rod-like. Bendy regions are segregated to the RNAPs in the nucleoid. Rod-like toroidal supercoiled regions segregate to between polymerases in the nucleoid. The walk of a hybrid helix to the RNAP is a random walk as diffusion must be in a phase. Different lengths of hybrid helix have different average time periods to do this random walk. When one of these structures gets to the RNAP initiation occurs fast—all that is required is rotation of the RNAP. The formation of these hybrid helices is a function of the DNA sequence, the DNA toroidal supercoil density and the rNTP concentrations. The equilibrium between DNA helix and single strand DNA/rNTP monomer hybrid helices is adjusted by the rNTP monomer pool levels establishing the differential pattern of transcription for many genes. For this model of transcription regulation to work requires all DNA in the nucleoid to be equivalent. What equivalence of all DNA in the nucleoid means is there are no barriers to the following equilibrium:

$$\text{Helix} \leftrightarrow \text{Helix} + \text{Supercoils} \leftrightarrow \text{Helix} + \text{Supercoils} + \text{Bubbles, kinks, hinges, random coils}$$

As we go from left to right in this equation the number of negative supercoils is increasing. rNTP monomers pull this equilibrium to the right. This equilibrium also defines a role for the known transcription regulatory proteins. Transcription regulatory proteins can "use" the Thermal Code, as also can regulatory proteins in the cytoplasm.

Hybrid helices are a small fraction of the DNA sequence—probably around 1% of the DNA sequence. The differential transcription pattern is determined by the relative rNTP monomer concentrations and the DNA sequences they interact with. I call this the Thermal Code model because it runs on thermal energy.

The information channel width can be defined as the number of words in the code. This depends in a simple way on the number of monomers in a hybrid helix. If the longest hybrid helix is 12 bases there are 4^{12} words in the code. Because there are probably 3−12 monomers in a hybrid helix this mechanism provides the coding possibility for exquisite quantitative control of global and differential transcription.

Hybrid Helices in Replication

I believe hybrid helices also occur in replication. It seems probable that ahead of the DNA polymerase are hybrid helices of dNTP monomers/DNA single strands stabilized by proteins. This explains the high synthesis rate of ~ 1000 bases/second. Also in the semidiscontinuous replication model the RNA primers on the lagging strand would seem to indicate the prior existence of rNTP/DNA single strand hybrid helices.

The First Several rNTP Monomers Synthesized Mechanism for the Thermal Code

The second mechanism I am proposing for the regulatory chemical events in transcription is that the rate for the first several monomers, two or three or four, or more of the rNTP monomers being polymerized into RNA determine the rate of transcription for that gene. The rates for these first, second or third, or more covalent bonds to be made in the phosphodiester backbone are determined by the concentration of the respective rNTP monomers involved. This mechanism is guaranteed to exist but not guaranteed to be the critical regulatory event for transcription regulation. I believe this Thermal Code mechanism is also a regulator of transcription, along with the hybrid helices Thermal Code mechanism.

In E. coli on the sense strand the first nucleotide transcribed, the initiation site, the +1 position, is usually a purine, with A(51%) and G(42%). The −1 position is a C(55%) and the +2 a T(48%).

Summary of the Two Thermal Code Mechanisms

I believe through setting the global and differential transcription rates as described the rate of initiation for different genes extends from initiation every ~ 1 second to initiation every $10^{3\text{-}4}$ seconds.

Critical Role of the Pribnow Box

The Pribnow box plays four critical mechanistic roles in my models:

1. Location of the denaturation bubbles, kinks, hinges, random coil sequences that diffuse to the RNAPs.
2. Location of the rNTP monomer/DNA single strand hybrid helices.
3. Location where the RNAPs get into the helix.
4. Establishes a clean initiation site.

These key roles are exemplified and corroborated by the variation found in the different Pribnow sequences for different genes, that variation giving rise to different frequencies of transcription. In E. coli, for 298

genes, on the sense strand the following is the percentage of agreement with the consensus sequence TATAAT:

T(77%)A(76%)T(60%)A(61%)A(56%)T(82%)

The ribosomal genes have very high agreement with the all A and T consensus sequence, which being all A and T would have the lowest melting points, and are the highest frequency transcribed genes, again consistent with my models.

I believe the upstream −35 homology is a rigid, rod-like DNA helix making an edge to a bendy hybrid helix. This is required to make the collision of the RNAP with the Pribnow box centered bubble, kink, hinge, random coil the determinant event for transcription initiation.

I have different roles for the Pribnow homology between in vitro and in vivo. In vitro, I believe, the Pribnow region is the region of low melting temperature, T_M, because A and T rich where the strands separate and RNAP gets into the helix between the strands, forming the open complex. In vivo, I believe, the RNAP gets into the helix centered on the Pribnow box and that the Pribnow box also determines a well-defined start base, the +1 initiation site.

Relative Role of the Constant Code and the Thermal Code

The Constant Code and Thermal Code are co-linear—they exist on top of each other in roughly the same places on the DNA sequence. RNA coding regions not only reject initiation in the absence of rNTP monomers but also in the presence of rNTP monomers. Promoters function in vitro forming the open complex in the absence of monomers. In vitro promoters have strengths based on the Constant Code alone. The quantitative relation for the relative roles played by the Constant Code and Thermal Code is of great interest.

The Division of Labor of rNTPs

There is a pervasive role of nucleic acid monomers in recognition and energetics in cells. rNTP monomers are high energy and can be recognized and dock in a precise way with proteins and do work either phosphorylating other biomolecules or transferring energy to other biomolecules, or forming polymers and phosphate. Monomers regulate the activity of many proteins in addition to being a substrate. For the Thermal Code to work the monomer pool levels must encode the state of the cell. The division of labor among monomers in the cell accomplishes this—UTP used for cell wall, CTP used for cell membrane, GTP used for protein synthesis, and ATP the all-purpose cytoplasm energy currency. A role for modified monomers such as cAMP, ppGpp and ppGppp beyond their currently understood roles can easily be imagined. Modification can affect the partitioning between cytoplasm and nucleoid of monomers, or affect the formation and diffusion of hybrid helices, or interact with transcription regulatory proteins, or interact directly with the RNAP/DNA binary complex in transcription initiation or with the RNAP/DNA/RNA ternary complex in elongation.

Separating Monomers from the Cytoplasm Requires Energy

Work is required to take rNTP, dNTP and modified monomers out of solution and have in higher concentration for polymer synthesis and regulatory roles. The Nucleic Acid Mainframe does this work in the minimum energy way by distributing that work over the surface of the nucleoid. dNTP monomers are separated from the cytoplasm and available at higher concentration for DNA synthesis and this in part accounts for the high rate of DNA synthesis of ~ 1000 bases/second. rNTP monomers are also separated from the cytoplasm in a minimum energy way and are available in higher concentration for both RNA

synthesis, that is not so fast at ~ 45 bases/second, and for their regulatory roles described in my Thermal Code mechanisms. Is there a regulatory role for rNDPs and rNMPs? I don't know the answer to that question. rNDPs and rNMPs are created in the cytoplasm and converted to rNTPs in the cytoplasm. Do they partition into the nucleoid or do only rNTPs have access to the nucleoid? I don't know but suspect only rNTPs have access to the nucleoid. rNTPs partition to the nucleoid through interactions with DNA, both in the major groove and with single strand DNA, and can also interact with other rNTP monomers and with nucleoid associated proteins. Mg^{++} partitions into the nucleoid through electrostatic interactions with the negative phosphodiester backbones of the DNA and I believe rNTP monomers also interact with the Mg^{++} in the nucleoid, as if the Mg^{++} "lubricates" partitioning of rNTPs to the nucleoid.

Transcription Terminators

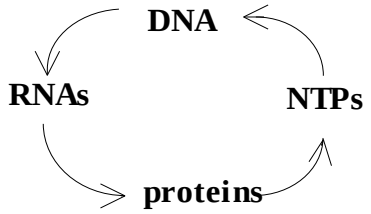
I have nothing to say about transcription terminators other than that they must be compatible with the Nucleic Acid Mainframe and Thermal Code models.

Phage

Bacterial viruses play the game of the Nucleic Acid Mainframe and the Thermal Code. I once described bacterial viruses as doing a "mathematical walk" through "concentration space".

Concluding Remarks

With the Nucleic Acid Mainframe and the Thermal Code the central dogma becomes circular rather than linear—NTP monomers feeding back on DNA to determine the RNA synthesis profile:



The Turing machine model is important to understanding the cell as an information processing machine. Turing machines in biology may seem a little esoteric but are very simple. Answering the questions how is DNA searched and how is the transcription pattern arrived at are the Turing machine description. Only small molecules like rNTP monomers can make the state of the cell available throughout the nucleoid. The entire DNA is continuously searched through interaction with monomers. Continuous search of the entire DNA is the most powerful Turing machine model of DNA possible. A large channel width—number of distinct words—also is essential to be a powerful Turing machine model. If transcription initiation, in the absence of regulatory proteins, is only regulated by the Constant Code the cell is a dumb Turing machine. An example of the type of calculations the cell Turing machine must do is to calculate the surface area to volume ratio given the size and shape of the cell and to calculate linear combinations of this ratio. In molecular terms, an example of such a calculation is how many lipids are needed for the membrane or how many of a membrane protein are needed for a given growth rate and size cell. Determining the number and timing of each protein synthesized, the number and timing of each stable RNA synthesized and the timing of the DNA replication decision, is the output of the cell Turing machine. The input is the state of the cell encoded in the rNTP concentration profile that evolves over the cell cycle and reflects the chemical environment the cell is growing in. The Nucleic Acid Mainframe and Thermal Code provide for information definition, transport, and processing in the cell—the basic smart Turing machine architecture.

The Thermal Code drives the binary fission growth cycle of prokaryotes.

The Thermal Code is not available in current in vitro systems. The nucleoid is a highly concentrated and exquisitely balanced system only found in vivo. The concentration of biomolecules is very much higher in vivo than in vitro. It may be difficult or impossible to get the Thermal Code to work in vitro, then again, starting with my qualitative model, it may be possible.

There are important molecular evolution and origin of life implications of the Thermal Code model. The Thermal Code must have preceded the genetic code. Before proteins are coded for in DNA transcription must have been occurring and a transcription code must have existed. The Thermal Code model allows for transcription to occur with primitive catalysts, possibly at first inorganic ion complexes as crude RNAPs, and later with coded for ribozyme "RNAPs". There no doubt were many ribozymes coded for in primitive cells, with primitive metabolisms, long before proteins arose. Of course, given the complexity of a cell, when thinking about the early evolutionary pathway it is always a bit difficult to understand how such a system could spontaneously arise.

To elucidate the Thermal Code at a fully quantitative level is a difficult task. A molecular simulation and kinetics analysis required to get the Thermal Code out will be difficult but possible starting out with my model.

There are important implications of the Thermal Code model to eukaryotic biology. At the root of cancer and developmental biology is the cell cycle. It seems likely at the root of the cell cycle in eukaryotes is an rNTP monomer based code. It even appears the rapid evolutionary creation of many eukaryotic organisms was propelled forward by the Thermal Code. The Nucleic Acid Mainframe and the RNA Processing Zone appear to be important in the nucleus of eukaryotic cells.