Applications of Differential Geometry to Molecular Genetics

A. M. FINDLEY*, S. P. McGLYNN† and G. L. FINDLEY*†

*Department of Chemistry, New York University, New York, NY 10003
†Department of Chemistry, Louisiana State University, Baton Rouge, LA 70803

Work supported by the U.S. Department of Energy.
†NYU Scheuer Presidential Fellow.

ABSTRACT: A mathematical formalism is presented in which changes in information content of an evolving DNA (deoxyribonucleic acid) molecule may be described. The basic construct is a 65-dimensional differentiable manifold (the *informational space-time manifold*) in a coordinate structure such that the manifold points represent (i) the number of each codon type in a DNA molecule, and (ii) the evolutionary time of that DNA. It is shown that this manifold cannot be Euclidean but must be taken, at least conditionally, to be Riemannian. Evolutionary motions in the informational space-time manifold are initially postulated to be geodesics, and evolutionary equations-of-motion are elaborated. These equations are governed by an evolutionary field which is produced by the intrinsic structure of the manifold. The concept of genetic cosmology is introduced, and a manifold in which the evolutionary field is weak and depends only upon the evolutionary time is investigated. The nature of empirical input into genetic cosmology is discussed.

1. INTRODUCTION

A. Prefatory Comments

At the molecular level, the *information content** of a gene is dictated by the linear arrangement of the nucleic acid bases in a DNA (deoxyribonucleic acid) molecule. Since the inception of this tenet (Watson and Crick, 1953), extensive chemical and biological studies have enhanced our knowledge of the fundamental processes that comprise *molecular genetics*.†

As a discipline, molecular genetics examines the physico-chemical basis for heredity. At the cornerstone of this field is the so-called "central dogma": that is, that DNA *replication*, DNA *transcription* into mRNA (messenger ribonucleic acid), and mRNA *translation* into protein serve as the molecular basis for the hereditary process. A vast amount of experimental work has gone into the explication of the enzymic reactions that constitute these processes and what has emerged is an overall picture that appears to be remarkably constant among all living systems.

Perhaps one of the most exciting prospects in molecular genetics is the possibility of investigating biological evolution through changes in the linear base sequences of DNA molecules. Such analyses have not been practicable because of the lack of detailed DNA sequence data. Analogous studies at the protein level, on the other hand, have benefited from the availability of protein sequence data (Dayhoff, 1972) and are well established (Fitch and Margoliash, 1967; Wu et al., 1974; Margoliash, 1975). However, with the advent of rapid DNA sequencing techniques (Sanger et al., 1977), it is certain that compendia of DNA sequences, from various parts of the genome in diverse organisms, will soon become available.

B. Description of Intent

We present a mathematical formalism in which changes in information content of a DNA molecule undergoing evolution may be described. We eschew any discussion of the complex molecular interactions that underlie the actual physical nature of an evolutionary system. Thus, our mathematical theory is predicated on the biological concepts of information storage, retrieval and processing systems and, in this sense, is outside the scope of traditional biophysics.

Our prime motivation is to formulate a theory in which meaningful questions may be posed concerning the relevance of symmetry concepts in biological evolution. Questions pertaining to biological symmetry are critically important because the detection of such inherent symmetries is comparable to the discernment of biological laws (Findley and McGlynn, 1979a, 1980; Findley et al., 1982a,b; Findley and Findley, 1982, 1984a, b). Thus, a realization of a part of biology as a mathematics is, in essence, the beginning of research which may ultimately lead to the explication of biological laws (Findley and McGlynn, 1979b, 1981); and this we take as sufficient justification for our viewpoint.

Of course, we do not fulfill all of these goals. What we do achieve, however, is a demonstration of the utility and feasibility of our approach to problems in evolution theory.

C. Précis of Formalism

The genesis of our development resides in a 65-dimensional differentiable manifold in a coordinate structure such that each manifold point represents (i) the number of each DNA codon type in a DNA molecule, whether such molecule is physically realizable or not and (ii) the evolutionary time of the molecule. This manifold is termed the *informational space-time manifold*, and its construction and a biological interpretation are the topics of Section 2. Specific considerations of the general biological problem lead one to conclude that this manifold cannot be Euclidean but must be taken, at least conditionally, to be Riemannian.

In Section 3, curves in the informational space-time manifold are interpreted as representing the evolutionary progress of DNA molecules. A provisional postulate is made that evolutionary motions in the informational space-time manifold are geodesics. It is shown, then, that the intrinsic structure of the manifold determines a biological *evolutionary field*, and *evolutionary equations-of-motion* are elaborated. The essential result is that the solution to evolutionary questions formulated at the DNA level resides, in principle, in the intrinsic structure of the informational space-time manifold: that is, in the knowledge of the biologically correct *genetic cosmology*.

The intrinsic structure of the manifold is, of course, determined by the fundamental tensor (cf. Section 3) and, in Section 4, we investigate a genetic cosmology in which the fundamental tensor is diagonal and a function only of the evolutionary time. In addition, the evolutionary equations-of-motion for a weak evolutionary field, which is solely evolutionary time-dependent, are given and the nature of future empirical input into genetic cosmology is discussed.

2. THE INFORMATIONAL SPACE-TIME MANIFOLD

A. Nature of the Manifold

Consider the set of DNA bases $B_\alpha \equiv (T,C,A,G)$, where T ≡ thymine, C ≡ cytosine, A ≡ adenine and G ≡ guanine. In terms of B_α, the set of 64 DNA codons C_α is prescribed by

$$C_\alpha \equiv B_\alpha \times B_\alpha \times B_\alpha \qquad (1)$$

where "x" denotes the Cartesian product. If we (i) consider each element of C_α to be an abstract vector; and (ii) consider the set of vectors C_α to be linearly independent over the real field, then C_α may be taken to be a basis for R^{64}, where R is the set of real numbers. We term the differentiable manifold for which C_α provides a class of C^∞-equivalent coordinate structures, the *informational space manifold* D.[§] Thus, a particular coordinate structure of D, the X structure, say, is specified by the set of 64-tuples of the 64 independent variables $X \equiv (x^1,\ldots,x^{64})$ over the real field. In terms of the tensor calculus, the components x^i, $i = 1,\ldots,64$, form a contravariant vector in the X structure.

In a particular coordinate structure of D, each point for which all of the x^i are non-negative integers is interpreted as representing the codon population of a single-stranded DNA (or DNA segment)[¶] : That is, at the point D, the coordinate x^i of D is the number of codons of type $c_i \in C_\alpha$ of the single-stranded DNA represented by D. Three amplifications of this interpretation are required.

--- For a given single-stranded DNA molecule (hereafter referred to as simply a DNA), the representative point of this DNA in $D[X]$ (a particular coordinate structure of D) does not contain information specifying the actual linear arrangement of codons in the DNA.[‖] Indeed, this consideration motivates our development of the informational space-time manifold (*vide infra*).

--- There exist points in $D[X]$ that do not represent physically realizable DNAs, namely, those points having at least one x^i which is not a non-negative integer. Two cases must be discussed: (i) those points of $D[X]$ having at least one x^i which is non-negative but non-integer; and (ii) those points of $D[X]$ having at least one x^i which is negative.

Case (i) points represent nonphysically realizable (virtual) DNAs. However, such points have an immediate biological interpretation: They are considered to represent an average of the coordinates of biologically related DNA molecules. Thus, case (i) points extend the possibility of providing a statistical interpretation of $D[X]$. In fact, we have two alternate biological interpretations of $D[X]$. If we take $D[X]$ to represent individual molecules, then the case (i) points are interpreted as virtual DNAs. On the other hand, if we take $D[X]$ to represent average DNAs (an average of the coordinates of biologically related DNAs over an entire species, say), then each non-negative point is physically realizable in a statistical sense. The dynamical formalism which we present below is the same regardless of which biological interpretation we choose and, hence, we will expand the term "DNA" to include either a specific single-stranded DNA molecule, or some appropriate average of single-stranded DNA molecules.

Case (ii) points require additional consideration, however. In physical spaces, the actual values of the coordinates have no absolute meaning since it is to changes in coordinates that physical laws apply. In $D[X]$, on the other hand, each point appears to have absolute meaning as the mathematical representation of a DNA. Thus, it would seem that we must reserve biological meaning for only the non-negative part of $D[X]$. We do so only in the sense that such points (those with no negative coordinates x^i) have immediate biological interpretation. We do not exclude the possibility, however, of the existence of biological laws which might eventually indicate that only coordinate differences have absolute meaning.

--- Since the points of $D[X]$ represent

information that ultimately derives from molecular configurations, then that subset of $D[X]$ which has biological meaning must be bounded. In an absolute sense, such boundedness is directly related to the finite amount of matter available to an evolving system.

Clearly, our initial interpretation of a point of $D[X]$ as representing the information content of a DNA is inappropriate since such a point in $D[X]$ identifies two DNAs that differ only in the linear arrangement of codons. In what follows, we remove this information loss with respect to evolution.

Our initial step is to modify the manifold D to produce a manifold that is appropriate for dynamical (evolutionary) questions. This is accomplished by increasing the dimension of D by one to account for evolutionary time. The resulting differentiable manifold is termed the *informational space-time manifold* M. Thus, in a particular coordinate structure, say χ, $M[\chi]$ is specified by the set of 65-tuples of the 65 independent variables $\chi \equiv (x^0, x^1,\ldots,x^{64})$ over the real field. $x^0 \equiv t$ is the evolutionary time coordinate and, as before, x^i, $i = 1,\ldots,64$, are the codon coordinates of $D[X]$. Again, in terms of the tensor calculus, the components x^μ, $\mu = 0,1,\ldots,64$, form a contravariant vector of $M[\chi]$.[**]

B. Evolutionary Motions

Consider two DNA molecules DNA_1 and DNA_2, which are connected by evolution: That is, let DNA_2 result from the evolution of DNA_1. DNA_1 and DNA_2 are represented by the points M_1 and M_2, respectively, in $M[\chi]$. Let the coordinates of M_1 be x^μ and those of M_2 by x'^μ, where $x^0 < x'^0$. The points M_1 and M_2 may be connected by infinitely many curves in $M[\chi]$, and each of these curves will be termed an *evolutionary motion*.

If we take M to be Euclidean, then the distance between M_1 and M_2 is well defined as the length of the Euclidean straight line connecting M_1 and M_2. An ambiguity arises at this point, however. If DNA_1 and DNA_2 differ only in the linear order of their codons, then $x^i = x'^i$. Hence, the distance between M_1 and M_2 is

$$d_{12} = \left[\sum_\mu (x'^\mu - x^\mu)^2\right]^{1/2} = (x'^0 - x^0) \equiv \Delta t \quad (2)$$

If, on the other hand, DNA_1 and DNA_2 do not differ at all (that is, if no biological evolution has occurred and DNA_1 has simply propagated linearly in time), then we still have $x^i = x'^i$. Furthermore, the distance between M_1 and M_2 is exactly the same as that specified by Eq. 2. We conclude that we cannot distinguish, metrically, between (i) the case in which evolution proceeds only through a permutation of the linear order of codons in a DNA (*permutational evolution*)[*†], and (ii) the case in which no biological evolution has occurred. Thus, the choice of a Euclidean structure for M is unsatisfactory and we now consider more general manifold structures.

5. EVOLUTIONARY MOTIONS TREATED AS GEODESICS

A. Geodesics

In terms of local differential geometry (Eisenhart, 1926), the differential element of length ds in $M[\chi]$ is given by [*‡]

FOOTNOTES IN TEXT

[*]Our use of the term "information content" is consonant with the terminology of molecular genetics in which DNA, RNA, and protein are referred to as "informational macromolecules". However, by this usage we intend no reference to information theory.

[†]It is beyond the scope of this article to synopsize molecular genetics. The interested reader should see larger reviews (Watson, 1976; Stent and Calendar, 1978; Birge, 1981).

[‡]We assume a familiarity with local differential geometry at the level of Eisenhart (1926).

[§]For the sake of completeness, we note that C_α is chosen as the basis because of its convenience in defining transcription and translation operators. This topic is not covered in this work but the interested reader is referred to Findley (1978).

[‖]A generalization of the theory to double-stranded DNAs will be found in Findley (1978). For our purposes, however, it is sufficient to consider only single-stranded DNAs since, from a given strand, it is always possible to generate the complementary strand.

[¶]It is possible to remove this difficulty via a formulation of a discrete space over a finite field (Findley, 1978). This formulation, however, does not permit generalization to a dynamical space in a consistent fashion.

[**]Greek indices run over the index set $(0,1,\ldots,64)$, while Latin indices run over the index set $(1,\ldots,64)$.

[*†] The concept of permutational evolution is also discussed by Findley and McGlynn (1979b, 1981).

[*‡]We adopt the summation convention: that is, if an index is repeated in one term, once in a contravariant position and once in a covariant position, then that index is summed over.

[*§] The reader should note that when $g_{\mu\nu} = 0$ for $\mu \neq \nu$, it follows that $g^{\mu\mu} = (g_{\mu\mu})^{-1}$.

[*¶] A likely candidate would be DNA sequences coding for various cytochromes c since extensive correlations exist for these moieties (Fitch and Margoliash, 1967; Wu et al., 1974; Margoliash, 1975).

[*‖]A compact set of real numbers is both closed and bounded.

$$ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu \qquad (3)$$

where $g_{\mu\nu} = g_{\mu\nu}(x^0, x^1, \ldots, x^{64})$, and where $g_{\mu\nu}$ is a symmetric, covariant, order-two tensor having non-zero determinant. The invariant ds^2 is termed the *metric*, and a differentiable manifold having a positive-definite metric is said to be *Riemannian*. The tensor $g_{\mu\nu}$ is referred to as the *fundamental tensor* and is totally determinative of the intrinsic structure of M (*vide infra*).

Consider two points M_1 and M_2 in $M[\chi]$ and a curve C connecting these two points. If the coordinates of the points of C are given as functions of a general parameter τ, such that $x^\mu(\tau_1)$ and $x^\mu(\tau_2)$ are the coordinates of M_1 and M_2 respectively, then we may define the integral

$$s = \int_{\tau_1}^{\tau_2} \left[g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \right]^{1/2} d\tau \qquad (4)$$

where s is the arc length of the curve C between the points M_1 and M_2. If s is stationary (that is, if, upon holding the endpoints of C constant, the first-order variation of s vanishes), then C is a *geodesic*, which is simply the generalization of the Euclidean straight line to curved spaces. By the techniques of the calculus of variations (Eisenhart, 1926), the stationarity of C implies

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{ds} \frac{dx^\sigma}{ds} = 0 \qquad (5)$$

where the parameter τ has been taken to be the arc length s, and where the *Christoffel symbols* $\Gamma^\mu_{\nu\sigma}$ are defined as

$$\Gamma^\mu_{\nu\sigma} = g^{\mu\gamma}\, \Gamma_{\gamma\nu\sigma} \qquad (6)$$

and

$$\Gamma_{\gamma\nu\sigma} = \frac{1}{2}\left(g_{\gamma\nu,\sigma} + g_{\gamma\sigma,\nu} + g_{\nu\sigma,\gamma}\right) . \qquad (7)$$

In Eq. 7 we have used the notation

$$g_{\gamma\nu,\sigma} = \frac{\partial g_{\gamma\nu}}{\partial x^\sigma} . \qquad (8)$$

The extremals of Eq. 4, where the parameter τ is again taken to be the arc length s, are the integral curves of the 65 ordinary differential equations given in Eq. 5. These are the geodesic curves, and they satisfy the condition that, anywhere along the curve (Eisenhart, 1926),

$$g_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 1 . \qquad (9)$$

B. Evolutionary Equations-of-Motion

In a manifold that does not possess a positive-definite metric, there exist curves of zero length. Since such a curve in $M[\chi]$ appears, at present, to have no biological meaning, we take $M[\chi]$ to be Riemannian. In addition, we make the following postulate: *Evolutionary motions in the informational space-time manifold are geodesics.*

This postulate selects certain curves, the geodesics, and distinguishes these as being entirely descriptive of evolution. Hence, the evolutionary motions satisfy Eq. 5. The components $\Gamma^\mu_{\nu\sigma}$ are determinative of the nature of an evolutionary motion and, therefore, may be thought of as comprising an *evolutionary field* on the informational space-time manifold. In this sense, then, Eq. 5 represents evolutionary equations-of-motion. Since the $\Gamma^\mu_{\nu\sigma}$ determine the curvature of the manifold (Eisenhart, 1926), it follows that evolution may be viewed as resulting from the curvature of the informational space-time manifold.

The identification of geodesics with evolutionary motions is provisional. In fact, this postulate may be generalized to include the case in which not all of evolution is governed by the curvature of the informational space-time manifold (Findley, 1978). Such a generalization requires, in a straightforward manner, the introduction of the concept of an *extrinsic* evolutionary field; this extension is addressed elsewhere (Findley et al., 1986).

4. GENETIC COSMOLOGY

The fundamental result of the previous section is that the solution to evolutionary questions formulated at the DNA level resides, in principle, in the intrinsic structure of the informational space-time manifold: that is, in the biologically correct *genetic cosmology*.

In examining the manifold structure, we discuss three cases: (i) the absence of an evolutionary field; (ii) the presence of a permutational evolutionary field; and (iii) an (incomplete) model genetic cosmology in which the evolutionary field is a function only of the evolutionary time. A comparison of the results of (i) and (ii) resolves the information loss discussed in Section 2.A. Finally, we comment briefly on empirical input into genetic cosmology.

A. Absence of an Evolutionary Field

The rectilinear propagation, in evolutionary time, of a DNA which is not evolving is specified by the following two conditions:

C-1. The $g_{\mu\nu}$ are constant for all μ, ν.

C-2. $v^i \equiv dx^i/ds = 0$ for all i.

C-1 ensures that the manifold be rectilinear (flat) and, since we are considering only Riemannian metrics, the manifold must be

Euclidean. The second condition ensures that all of the informational space coordinates be constant along an evolutionary motion.

For an evolutionary motion, application of C-2 to Eq. 9 yields

$$g_{00}v^{0^2} = 1 \tag{10}$$

or,

$$v^0 = g_{00}^{-1/2}. \tag{11}$$

Thus, the evolutionary equation-of-motion is

$$\frac{dv^0}{ds} = \frac{d}{ds}\ (g_{00}^{-1/2}) = 0 \tag{12}$$

where the second equality follows from C-1.

Now, from Eq. 10, we have

$$ds = g_{00}^{1/2}\ dx^0 \equiv g_{00}^{1/2} dt \tag{13}$$

Because of C-1, however, we may choose $g_{00} = 1$. The arc length of the evolutionary motion from point M_1 to point M_2 is then given by

$$s = \int_{t_1}^{t_2} dt = \Delta t \tag{14}$$

which is, of course, the same result as that of Section 2.B. Thus, a DNA which is not evolving is characterized by 64 coordinates x^i which are constant and coordinate $x^0 \equiv t$ which is propagating in a linear manner.

B. Presence of a Permutational Evolutionary Field

We next consider a DNA which is evolving so that only the linear order of its codon changes: that is, the informational space coordinates are constant along an evolutionary motion. Such evolution is specified by the following two conditions:

C'-1. $$g_{\mu\nu} = \begin{cases} g_{\mu\mu}(t) & \text{for } \mu = \nu \\ 0 & \text{for } \mu \neq \nu \end{cases}$$

C'-2. $v^i = 0$ for all i.

For an evolutionary motion, application of C'-2 to Eq. 9 yields

$$g_{00}v^{0^2} = 1. \tag{15}$$

Clearly, the difference between Eqs. 15 and 10 is that g_{00} is a function of evolutionary time in Eq. 15, but constant in Eq. 10.

Rewriting Eq. 15 as

$$v^0 = g_{00}^{-1/2} \tag{16}$$

we find that the evolutionary equation-of-motion is

$$\frac{dv^0}{ds} = \frac{d}{ds}\ (g_{00}^{-1/2}) \quad . \tag{17}$$

However, Eq. 15 implies

$$ds = g_{00}^{1/2}\ dx^0 \equiv g_{00}^{1/2}\ dt \tag{18}$$

and, hence, we find

$$\frac{dv^0}{ds} = g_{00}^{-1/2}\ \frac{dv^0}{dt}\ ;$$

$$\frac{d}{ds}\ (g_{00}^{-1/2}) = g_{00}^{-1/2}\ (g_{00}^{-1/2})_{,0} \tag{19}$$

By substituting Eq. 19 into Eq. 17, the evolutionary equation-of-motion becomes

$$\frac{dv^0}{dt} = (g_{00}^{-1/2})_{,0} \quad . \tag{20}$$

Finally, from Eq. 18, the arc length of the evolutionary motion from the point M_1 to the point M_2 is given by

$$s = \int_{t_1}^{t_2} g_{00}^{1/2}(t)\ dt \tag{21}$$

where we have explicitly indicated the evolutionary time dependence of $(g_{00})^{1/2}$.

The content of Eq. 21 is elucidated by comparison with Eq. 14: for the case in which evolution is manifested through codon permutations only, the arc length of the evolutionary motion between two points in the informational space-time manifold is a *nonlinear* function of the evolutionary time coordinates of the points; in the absence of an evolutionary field, however, the arc length of an evolutionary motion between two points in the informational space-time manifold is a *linear* function of the evolutionary time coordinates of the points.

The choice of a Euclidean manifold structure inherently results in an information loss that identifies the distance between two points in $M[\ ]$ regardless of whether they represent the same DNA at different evolutionary times or two distinct DNAs, varying only in codon order, at different evolutionary times (cf. Section 2.B). From the above derivation, we conclude that a curved informational space-time manifold restores the information loss, with respect to evolution, inherent in the original Euclidean formulation.

C. An (Incomplete) Model Genetic Cosmology

We shall now consider a model genetic cosmology that incorporates the results of Sections 4.A-B as special cases. In addition, we make the assumption that the evolutionary field is weak (i.e., that biological evolution occurs slowly). Thus, the conditions are

C"-1. $$g_{\mu\nu} = \begin{cases} g_{\mu\nu}(t) & \text{for } \mu = \nu \\ 0 & \text{for } \mu \neq \nu \end{cases}$$

C"-2. $\lim_{q^{\mu\nu} \to} g_{\mu\nu} = c_{\mu\nu}$, where the $c_{\mu\nu}$ are constants for all μ.

C"-3. v^i is a small quantity of the first order with respect to v^0 for all i.

The model is incomplete in the sense that we never specify an exact functional form for the $g_{\mu\nu}$.

In view of C"-1, we find that the only non-zero $\Gamma^{\mu}_{\nu\sigma}$ are

$$\Gamma^{i}_{i0} = \frac{1}{2} g^{ii} g_{ii,0} \quad ; \quad \Gamma^{0}_{ii} = -\frac{1}{2} g^{00} g_{ii,0} \tag{22}$$

$$\Gamma^{0}_{00} = \frac{1}{2} g^{00} g_{00,0}$$

Substitution of Eq. 22 into Eq. 5 yields the evolutionary equations-of-motion

$$\frac{dv^i}{ds} = - g^{ii} g_{ii,0} v^i v^0 \tag{23a}$$

and

$$\frac{dv^0}{ds} = - \frac{1}{2} g^{00} \left\{ g_{00,0} v^{0^2} - g_{ii,0} v^{i^2} \right\} \tag{23b}$$

Imposing the condition C"-2, we see that the equations-of-motion become

$$\frac{dv^{\mu}}{ds} = 0 \tag{24}$$

in the limit of $t \to \infty$, thus regenerating the results of Section 4.A. For those evolutionary motions for which $v^i = 0$, the equations-of-motion become

$$\frac{dv^0}{dt} = (g_{00}^{-1/2})_{,0} \tag{25}$$

thus regenerating the results of Section 4.B.

Eqs. 23a,b may be simplified by applying C"-3. For an evolutionary motion, then, we find (cf. C"-1 and Eq. 9)

$$g_{00} v^{0^2} = 1 \tag{26}$$

where in Eq. 26 we have neglected terms which are of the second order in smallness. Applying C"-3 to Eqs. 23a,b and making use of Eq. 26, we find

$$\frac{dv^i}{dt} = - g^{ii} g_{ii,0} v^i \tag{27a}$$

and

$$\frac{dv^0}{dt} = (g_{00}^{-1/2})_{,0} \tag{27b}$$

where all second-order terms have been neglected.

Eqs. 27a,b are the evolutionary equations-of-motion for a weak evolutionary field which is a function only of the evolutionary time.

We may rewrite Eq. 27a as

$$\frac{dv^i}{dt} = (\ln g^{ii})_{,0} v^i \tag{28}$$

Remembering that $v^i \equiv dx^i/ds$ and integrating Eq. 28 twice over the interval [0,t], we find

$$x^i(t) = \int_0^t g_{00}^{1/2} \left\{ \int_0^{t'} g_{00}^{-1/2} (\ln g^{ii})_{,0} \frac{dx^i}{dt''} dt'' \right\} dt'$$

$$+ \left[g_{00}^{-1/2} \frac{dx^i}{dt} \right]_{t=0} \cdot t + [x^i]_{t=0} \tag{29}$$

Eq. 29 is a formal solution to the question of how the codon population of a DNA changes with evolutionary time, in the presence of a weak evolutionary field which is a function only of the evolutionary time.

Further investigations of Eq. 29 require a choice of the time functionality of the fundamental tensor components $g_{\mu\nu}$. Only then does the theory presented here become a complete model genetic cosmology.

D. Empirical Genetic Cosmology

The raw data for generating empirical gentic cosmologies, and for testing model genetic cosmologies, are collections of DNA sequences from a large number of different species. From such compilations one proceeds as follows:

--- Order the species, evolutionarily, using classical phylogenetic methods or presently available phylogenetic methods (Dayhoff, 1972; Fitch and Margoliash, 1967; Wu et al., 1974; Margoliash, 1975).

--- From each species, select a DNA sequence which codes for a protein having essentially the same function in all species.

The result of this process is the construction of a series of homologically ordered DNA sequences with an associated approximate evolutionary time scale. Such a series can be used to suggest, in a rough manner, the functional form of the fundamental tensor. Clearly, this will require much trial-and-error model building. However, the difficulty with this ansatz is the paucity of DNA sequences presently available. Detailed empirical research in genetic cosmology, therefore, must wait until such data become available.

5. DISCUSSION

This work, formal though it may be, represents a significant simplification: namely, attention is redirected from the complex physicochemical processes involved in evolution (as mediated through natural selection) to the totally geometric concept of an evolutionary field generated by the curvature of the

informational space-time manifold. Such a treatment is crucial to any construction of a theory in which the concept of biological symmetry is stripped of its imprecision. In view of our ability to represent the change in information content of an evolving DNA, we expect that this development will facilitate future analyses of the nature of symmetry in biological evolution formulated at the level of molecular genetics. Our intention in this paper was to show that these concepts are sufficient to account for evolution; the demonstration of their necessity is under study (Findley et al., 1986).

Although quantitative research into the nature of the biologically correct genetic cosmology is currently impracticable, there still exists a wealth of classical evolutionary ideas which may lead to qualitative statements concerning the manifold structure. For example, biological evolution is generally thought to be *divergent* (Beyer et al., 1974). At the molecular level, this means that, starting from a single DNA, two lines of evolutionary descent never terminate, at the same evolutionary time, in a DNA which is the same for both lines of descent. In terms of the theory presented here, divergent evolution merely signifies that two evolutionary motions emanating from the same point never cross. In turn, this implies that the biologically meaningful subset of the informational space-time manifold cannot be compact.[11] However, in terms of the discussion of Section 2.A, we conclude that the subset of the manifold which has biological meaning cannot be closed.

Suppose, on the other hand, that these same two lines of descent terminate, at the identical evolutionary time, in DNAs which differ only in codon order. Such an event represents a crossing of two evolutionary motions emanating form the same point. However, from an evolutionary point of view, divergence has not been violated. Thus, if the manifold structure is such that evolution is divergent, then the above termination of the two lines of descent is not possible. This, of course, is a prediction which must be tested by experiment.

As another example, we cite the dependence of a gene's mutation rate upon the nature of the particular gene in question (Dobzhansky et al., 1977). Such a dependence intimates that the model presented in Section 4.C is, perhaps, too simple-minded since it precludes any dependence of the fundamental tensor upon the codon coordinates.

These brief remarks should indicate the necessity of continued, qualitative research in genetic cosmology. In the end, however, the fundamental test of our viewpoint must await detailed empirical analyses.

REFERENCES

Beyer, W.A.; Stein, M.L.; Smith, T.F.; Ulam, S.A. 1974. *Math. Biosci. 19*, 9-25.

Birge, E.A. 1981. *Bacterial and Bacteriophage Genetics, An Introduction*, Springer-Verlag, New York.

Dayhoff, M.O. 1972. *Atlas of Protein Sequence and Structure*, Natl. Biomed. Res. Found., Washington, DC.

Dobzhansky, T.; Ayala, F.J.; Stebbins, G.L.; Valentine, J.W. 1977. *Evolution*, Freeman, San Francisco, pp. 68-72.

Eisenhart, L.P. 1926. *Riemannian Geometry*, Princeton Univ. Press, Princeton, NJ.

Findley, A.M.; Findley, G.L. 1982. *Int. J. Quantum Chem., Quantum Biol. Symp. 9*, 59-63.

Findley, A.M.; Findley, G.L. 1984a. *Int. J. Quantum Chem., Quantum Biol. Symp. 11*, 109-116.

Findley, A.M.; Findley, G.L. 1984b. In: *Colloquium in Biological Sciences*, New York Acad. Sci., NY, pp. 537-540.

Findley, A.M.; Findley, G.L.; McGlynn, S.P. 1982a. *J. Theor. Biol. 97*, 299-318.

Findley, A.M.; McGlynn, S.P.; Findley, G.L. 1986. *The Geometry of Genetics*, Wiley-Interscience, New York, in press.

Findley, G.L. 1978. *A Riemannian-Geometric Realization of Molecular Genetics*, Ph.D. Thesis, Louisiana State University, Baton Rouge.

Findley, G.L.; McGlynn, S.P. 1979a. *Int. J. Quantum Chem., Quantum Biol. Symp. 6*, 313-327.

Findley, G.L.; McGlynn, S.P. 1979b. In: *Group Theoretical Methods in Physics*. (W. Beiglbock, A. Böhm and E. Takasugi, eds.), Springer-Verlag, Berlin, pp. 71-72.

Findley, G.L.; McGlynn, S.P. 1980. *Int. J. Quantum Chem., Quantum Biol. Symp. 7*, 277-281.

Findley, G.L.; McGlynn, S.P. 1981. *Int. J. Quantum Chem., Quantum Biol. Symp. 8*, 455-461.

Findley, G.L.; Findley, A.M.; McGlynn, S.P. 1982b. *Proc. Natl. Acad. Sci. USA 79*, 7061-7065.

Fitch, W.M.; Margoliash, E. 1967. *Science 155*, 279-284.

Margoliash, E. 1975. *Adv. Chem. Phys. 29*, 191-252.

Sanger, F.; Air, G.M.; Barrell, B.G.; Brown, N.L.; Coulson, A.R.; Fiddes, J.C.; Hutchinson, C.A., III; Slocombe, P.M.; Smith, M. 1977. *Nature 265*, 687-695.

Stent, G.S.; Calendar, R. 1978. *Molecular Genetics, An Introductory Narrative*, 2nd ed., Freeman, San Francisco.

Watson, J.D. 1976. *Molecular Biology of the Gene*, 3rd ed., Benjamin, Menlo Park, CA.

Watson, J.D.; Crick, F.H.C. 1953. *Nature 171*, 964-969.

Wu, T.T.; Fitch, W.M.; Margoliash, E. 1974. *Ann. Rev. Biochem. 43*, 539-566.