Journal of Theoretical Biology 401 (2016) 94–108

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# Nonlinear effects in evolution – an *ab initio* study: A model in which the classical theory of evolution occurs as a special case

Daryl G. Clerc *

*ArTek Product Development, Inc., P.O. Box 212, Red Bud, IL 62278, USA*

## HIGHLIGHTS

- Changes in phenotypic properties may be studied in terms of binding affinity.
- Nonlinear and linear contributions are equally significant.
- Organs progress with varying degrees of interdependence.
- Likelihood of successful mutation decreases with increasing chemical complexity.
- The model and classical theory diverge with increasing complexity of the organism.

## ARTICLE INFO

*Article history:*
Received 14 December 2014
Received in revised form
27 February 2016
Accepted 17 March 2016
Available online 28 March 2016

*Keywords:*
Chemical evolution
Evolutionary model

## ABSTRACT

An *ab initio* approach was used to study the molecular-level interactions that connect gene-mutation to changes in an organism's phenotype. The study provides new insights into the evolutionary process and presents a simplification whereby changes in phenotypic properties may be studied in terms of the binding affinities of the chemical interactions affected by mutation, rather than by correlation to the genes. The study also reports the role that nonlinear effects play in the progression of organs, and how those effects relate to the classical theory of evolution. Results indicate that the classical theory of evolution occurs as a special case within the *ab initio* model – a case having two attributes. The first attribute: proteins and promoter regions are not shared among organs. The second attribute: continuous limiting behavior exists in the physical properties of organs as well as in the binding affinity of the associated chemical interactions, with respect to displacements in the chemical properties of proteins and promoter regions induced by mutation. Outside of the special case, second-order coupling contributions are significant and nonlinear effects play an important role, a result corroborated by analyses of published activity levels in binding and transactivation assays. Further, gradations in the state of perfection of an organ may be small or large depending on the type of mutation, and not necessarily closely-separated as maintained by the classical theory. Results also indicate that organs progress with varying degrees of interdependence, the likelihood of successful mutation decreases with increasing complexity of the affected chemical system, and differences between the *ab initio* model and the classical theory increase with increasing complexity of the organism.



## 1. Introduction

The purpose of this study is to derive and evaluate an *ab initio* model of evolution in order to explore the associated chemical reactions at the molecular level, and thereby provide insight into the relationship between changes in an organism's genetic code and changes in its phenotype. The *ab initio* approach enables fundamental relationships between mutation and phenotype to be cast in a basic mathematical form, thereby enabling quantitative assessment of the associated nonlinear effects. The nonlinear effects arise from non-negligible contributions of interdependencies within the associated chemical processes.

Many types of interdependencies between organs and proteins are observed empirically. For example, the brain depends on many proteins such as Contactin-1 (*CNTN1*), α1-Syntrophin (*SNTA1*), γ1-Syntrophin (*SNTG1*), and γ2-Syntrophin (*SNTG2*), among others, where gene designation is parenthesized (UNIPROT). Some proteins are singly-located, such as γ1-Syntrophin (*SNTG1*) which is found only in the brain. Other proteins, such as α1-Syntrophin (*SNTA1*), sine oculis (*SIX1*) and the heart-type fatty acid binding

* Tel.: +1 618 340 9416.
*E-mail address:* proddev@artek1.com

http://dx.doi.org/10.1016/j.jtbi.2016.03.026

protein (*FABP3*) are found in many locations. Snta1 has high levels of expression in skeletal muscle and the heart, as well as low levels of expression in the brain, kidney, pancreas, liver and lungs (UNIPROT). In mice, Six1 is present in the olfactory system, statoacoustic system, pituitary gland, thymus, various muscle groups, tendons, and connective tissue (Laclef et al., 2003). Fabp3 is present in many cell types, such as heart, skeletal muscle, brain, kidney, lung and stomach, among others (Furuhashi and Hotamisligil, 2008). Another example of interdependence is the liver, which manufactures approximately 80% of all amino acids used in the body. Mutations that affect the *SNTA1*, *SIX1* or *FABP3* genes, or the liver, thus have the potential to affect multiple organs.

Activation of a protein can potentially be affected by mutations in several locations, given that regulatory processes are associated with multi-body interactions. For example, in order for transcription to commence in eukaryotes, a collection of transcription factor (TF) proteins attach to a gene's promoter region, often as homo-dimers or heterodimers, followed by the binding of RNA polymerase to the collection, to form a transcription initiation complex bound to the promoter site of the protein's gene. A protein's transcription is thus affected by mutations to associated TFs and co-activators that affect their ability to form the transcription initiation complex. Further, TFs possess a zone that binds to DNA, and a mutation to that zone is capable of changing the TF's ability to bind directly to the gene's activation site. Thus changes in a protein's regulation depend not only on mutations within the promoter segments of its own gene, but also on mutations to the zone of a TF that attaches to the promoter region, and on mutations within the coding and promoter segments in the genes of its other TFs.

Epistatic processes (gene-gene interactions) involve interdependencies as well, through the capability of one mutation to affect multiple organs and multiple chemical processes. In an epistatic process, mutation to one gene and its associated protein affects all related downstream organs, and overrides or masks the effects that mutations to downstream genes would ordinarily have. For example, a mutation in the liver could indirectly affect any organ that employs the amino acids that the liver generates. A gene associated with those particular organs would thereby be masked by the mutation to the liver. Another example is the mutation that causes albinism, which overrides the mutation that affects eye color.

This study differs fundamentally in its methodology from other mathematical theories of evolution, which have generally treated the process from a more macroscopic perspective. For example, quantitative genetics treats changes in phenotypic traits statistically, both within and between populations, as they relate to both genotypic and environmental factors (Walsh, 2001). One-locus and two-locus models involve the analysis of factors such as gamete characteristics and allele frequencies, among others, in order to evaluate the evolution of populations over time (Yanchukov, 2009). Coalescent theory involves tracing all alleles of a gene shared by members of a population to a single ancestral copy, known as the most recent common ancestor (Kingman, 1982a, 1982b, 1981). Evolutionary game theory involves analyzing the strategies employed in a competing population to survive and reproduce (Sandholm, 2007). Thus, this study differs in its approach compared to current theories, owing to its focus on chemical interactions at the molecular-level.

The term *ab initio*, whose literal meaning is "from the beginning", is used to describe this study given that it relates changes in phenotypic properties to molecular-level quantities which are derivable from first principles and calculable using the rigorous computational techniques of quantum chemistry. *Ab initio* methodology is therefore adhered to in this study, that is relying on established laws of nature without introducing assumptions. As such, scientific observations that have been thoroughly corroborated over long periods of time, to such an extent that they are currently considered self-evident, are considered established laws of nature. Examples include the dependence of an organ's physical properties on the proteins that comprise it, and the dependence of a protein's chemical composition on its encoding gene. External influences on the organism are not considered – such as climate, availability of food sources, group behavior, communication, and so forth. Focus is placed on the mutations that are transmitted to offspring, such as germ-line mutations in animals, and upon the associated chemical mechanisms whereby those mutations affect the phenotype of the offspring, at the level of first-principles.

## 2. Results

Results are reported in three sections. Initially, expressions that relate mutation of genes to changes in phenotype are derived. Second, the resultant expressions are compared to laboratory data to assess the level of coupling contributions. Lastly, the *ab initio* model is compared to the classical theory of evolution.

A number of terms are adopted in this study. The term "phenotype" refers to the physical and biochemical state of an organism, and "physical property" refers to macroscopic properties as well as biochemical properties. The term "protein" is used generically throughout to refer not only to a type of protein *per se*, but also to other types of compounds such as hormones, enzymes, and other protein-derivatives. The term "residue" denotes an amino acid residue in a protein, "nucleotide" denotes a nucleotide in the promoter region of a gene, and "entity" refers to either a residue or a nucleotide. Lastly, the term "organ" in a mathematical context refers to a collection of the physical properties of a biological system component.

## 3. Theory

In the following discussion, coding-region mutations which alter the composition of the corresponding protein are considered, such as a missense mutation whose transcription yields a different amino acid residue in the associated protein. Also considered are mutations to promoter regions, which alter production of the corresponding protein, such as a alteration of a nucleotide that affects binding to transcription factors. Definitions of all central variables are listed in the supplemental portion of this study (Supplementary material – Part S.1). All assertions are enumerated in this section, along with their associated mathematical expressions, as follows.

### 3.1. A mutation that alters the composition of a gene in its coding or promoter region, respectively affects the chemical properties of residues in its associated protein, or nucleotides in its promoter region

A mutation in the coding region of a gene that yields a change of composition in its corresponding protein, effectively replaces one type of residue with one or more different types of residues. Similarly, a mutation in the promoter region of a gene affects the chemical composition of the nucleotides there. As a result, the associated displacement of a chemical property at that location may be represented as follows,

$$\Delta\pi_{rm}^{(I,p\to q)} = \pi_{rm}^{(I)}|_q - \pi_{rm}^{(I)}|_p. \tag{1}$$

Here, $\pi_{rm}^{(I)}$ denotes the $m^{th}$ chemical property of the $r^{th}$ entity. The terms $\pi_{rm}^{(I)}|_p$ and $\pi_{rm}^{(I)}|_q$ denote a chemical property of the original and modified entity, respectively. Superscripted indices within parentheses denote parametric dependence. For example,