Rolling the Dice Twice:

Evolving Reconstructed Ancient Proteins in Extant Organisms

Betul Kacar

NASA Astrobiology Institute, Mountain View, CA 94035, U.S.A.
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, 02138, USA

Correspondence should be sent to kacar@g.harvard.edu

<A> Abstract

Scientists have access to artifacts of evolutionary history (namely, the fossil record and genomic sequences of living organisms) but they have limited means with which to infer the exact evolutionary events that occurred to produce today's living world. An intriguing question to arise from this historical limitation is whether the evolutionary paths of organisms are dominated by internal or external controlled processes (i.e., Life as a factory) or whether they are inherently random and subject to completely different outcomes if repeated under identical conditions (i.e., Life as a casino parlor). Two experimental approaches, ancestral sequence reconstruction and experimental evolution with microorganisms, can be used to recapitulate ancient adaptive pathways and provide valuable insights into the mutational steps that constitute an organism's genetic heritage. Ancestral sequence reconstruction follows a backwards-from-present-day strategy in which various ancestral forms of a modern gene or protein are reconstructed and then studied mechanistically. Experimental evolution, by contrast, follows a forward-from-present day strategy in which microbial populations are evolved in the laboratory under defined conditions in which their evolutionary paths may be closely monitored. Here I describe a novel hybrid of these two methods, in which synthetic components constructed from inferred ancestral gene or protein sequences are placed into the genomes of modern organisms that are then experimentally evolved. Through this system, we aim to establish the comparative study of ancient phenotypes as a novel, statistically rigorous methodology with which to explore the respective impacts of biophysics and chance in evolution within the scope of the Extended Synthesis.

Keywords: Synthetic biology, genome engineering, experimental evolution, paleomolecular biochemistry, elongation factor, historical contingency.

Living organisms are historical systems and their evolutionary history is one of the vital determinants of their capacity to respond to their environment. Historical contingency is a property of living systems that assigns

past events as important factors that shape a system's current state. In his book *Wonderful Life*, Stephen Jay Gould famously

posed a thought experiment to address whether current biota are the product of random evolutionary luck, or whether there

were other possible trajectories that could have taken place. Assigning historical contingency as a fundamental influence in

shaping evolutionary outcomes, Gould posited that if life's tape were rewound and replayed from various points in the

distant past, the resulting living world would be very different than it is now:

> I call this experiment "replaying life's tape." You press the rewind button and, making sure you
> thoroughly erase everything that actually happens, go back to any time and place in the past—say, to the
> seas of the Burgess Shale. Then let the tape run again and see if the repetition looks at all like the original.
> If each replay strongly resembles life's actual pathway, then we must conclude that what really happened
> pretty much had to occur. But suppose that the experimental versions all yield sensible results strikingly
> different from the actual history of life? What could we then say about the predictability of self-conscious
> intelligence? or of mammals? or of vertebrates? or of life on land? or simply of multicellular persistence
> for 600 million years? (Gould 1989)

Understanding to what degree historical contingency shapes evolutionary trajectories would be possible by

performing Gould's thought experiment. We cannot, of course, carry out this experiment at the global scale Gould

envisioned, however, methods of experimental evolution allow researchers to tackle some aspects of rewinding and

replaying evolution in controlled environments (Blount, Borland, and Lenski 2008; Fortuna et al. 2013; Losos 1994;

Travisano et al. 1995; Desjardins this volume). A recent advance in experimental biology also offers an approach that

permits the tape of life to be rewound for individual genes and proteins. Commonly referred as "ancestral sequence

reconstruction", or paleogenetics, this method integrates molecular phylogeny with experimental biology and *in vitro*

resurrection of inferred ancestral proteins (Figure 13.1; see also Dean and Thornton 2007). Here I present a novel approach

that merges ancestral sequence reconstruction with experimental evolution. In this method, my colleagues and I have

engineered a reconstructed ancestral gene directly into a modern microbial genome with the intention of observing the

immediate effects of this ancient protein in a modern bacterial context (Kacar and Gaucher 2012). This approach therefore

initially follows a backwards-from-present-day strategy in which we essentially recover a previous point on the tape of life,

30   reconstruct it (rewind), and then observe the interaction of this ancient protein with the present (replay). Before presenting

31   the details of our system, and where we would like to go with the *in vivo* resurrection of ancient genes, I will first detail how

32   ancestral sequence reconstruction provides insights into the past.

33

34   **<A> Ancestral Sequence Reconstruction: Using Phylogeny to Unravel Evolutionary History**

35   Almost half a century ago, Pauling and Zuckerkandl recognized the potential of phylogenetics to provide

36   information about the molecular history of life. In a pioneering paper titled "Chemical Paleogenetics: 'Molecular

37   Restoration' Studies of Extinct Forms of Life," they suggested a relatively straightforward methodology in which gene or

38   protein sequences obtained from existing organisms would be determined, aligned, and then used to construct phylogenetic

39   trees from which the ancestral states of modern genes and protein sequences could be inferred (Pauling and Zuckerkandl

40   1963). In 1990, Benner et al. realized the vision of Pauling and Zuckerkandl by reconstructing, synthesizing, and

41   characterizing an ancestral ribonuclease from an extinct bovid ruminant in the laboratory (Stackhouse et al. 1990).

42   Ancestral sequence reconstruction has since been used to examine the molecular history of many proteins,

43   allowing biologists to test various hypotheses from molecular evolutionary theory by tracing the evolutionary paths proteins

44   have taken to acquire their particular function (Benner, Sassi, and Gaucher 2007). This backward-from-today strategy has

45   also provided helpful insights about various the environmental conditions of the ancient Earth (Galtier, Tourasse, and Gouy

46   1999; Gaucher 2007; Ogawa and Shirai 2013), paved the way for emerging paradigms such as functional synthesis (Dean

47   and Thornton 2007), evolutionary synthetic biology (Cole and Gaucher 2011), and evolutionary biochemistry (Harms and

48   Thornton 2013). Moreover, it has led to the development of models of molecular evolution that have been used to guide

49   applications of evolutionary theory in medical and industrial settings (Kratzer et al. 2014; Risso et al. 2013).

50   Our work has focused on reconstructing the molecular past of Elongation Factor-Tu (EF-Tu) proteins. EF-Tu

51   proteins function to deliver aminoacylated-tRNA molecules into the A-site of the ribosome, and are thus essential

52   components of the cell (Czworkowski and Moore 1996). Moreover, EF-Tu proteins also interact with a variety of other

53   proteins functioning outside of translation machinery, and thus serve in activities outside of their role in the ribosome

54   (Defeu Soufo et al. 2010; Kacar and Gaucher 2013; Pieper et al. 2011).

55   Various ancestral sequences of bacterial EF-Tu, ranging from approximately 500 million to 3.6 billion years old,

56   have been previously inferred and reconstructed (Gaucher, Govindarajan, and Ganesh 2008). Useful to these studies was the

57   information that EF-Tu proteins exhibit a high correlation with it's organism's environmental temperature (Gromiha,

58   Oobatake, and Sarai 1999). For example, EF-Tus from thermophilic organisms (organisms living in high temperature

3

59   environments) exhibit high temperature tolerance and thermostability, while EF-Tus obtained from organisms that live in

60   moderate conditions (i.e. mesophiles) exhibit much lower temperature tolerance and stability. EF-Tus therefore appear to be

61   subject to strong selection to be optimized to their host's thermal environment.

62   The combination of EF-Tu's essential role in the cell, the strong selection acting on EF-Tu phenotype and the

63   availability of various phenotypically and genotypically altered ancient EF-Tu proteins created an ideal situation for us

64   rewind the molecular tape of life for one gene. A rather obvious question is whether a living organism could function if its

65   native EF-Tu was replaced with an ancient form? How and to what degree the reconstructed ancestral EF-Tu proteins would

66   interact with the components of the modern cellular machinery? Moreover, can we provide insights into the factors that

67   determine whether a reconstructed ancient EF-Tu, with very distinct genotypic and phenotypic properties, can interact with

68   modern components?

69   We therefore set out to first identify whether a recombinant bacteria, in which its endogenous EF-Tu is replaced

70   with an ancestral EF-Tu at the precise genomic location would result in a viable bacteria. First to assess was whether ancient

71   EF-Tu proteins would interact with the members of the translational machinery and thus perform their primary function. For

72   this, we measured the activity of various ancient EF-Tu proteins in an *in vitro* translation system. This system is composed

73   of translation machinery components recombinantly purified from modern *E. coli* bacteria, and individually presented in a

74   test tube, providing us control over which molecules are presented or omitted (Shimizu et al. 2001). We removed the

75   modern *E. coli* EF-Tu from this cell-free system, individually inserted ancient EF-Tu proteins one by one, and measured the

76   activity of the hybrid system by following fluorescent protein production by the translation machinery.

77   Our *in vitro* findings show that foreign EF-Tus (both ancient counterparts and modern homologs of *E. coli* EF-Tu)

78   exhibit function, albeit not equal to the *E. coli* EF-Tu, in a modern translation system where all the other components of the

79   translation system is obtained from *E. coli* (Zhou et al. 2012). Therefore while replacing the native, modern EF-Tu with an

80   ancient EF-Tu seems likely to put stress on *E. coli*, the recombinant organism is also likely to be viable. Although this is an

81   exciting observation, replacing the EF-Tu of a natural system with another EF-Tu protein that has co-adapted to a foreign

82   cellular system (be it a modern homolog or ancient) will always carry the risk that the interactions between the foreign

83   protein and its ancillary cellular partners may be damaged to the point of complete dysfunction (Figure 13.1). Such a

84   possibility restricts our ability to move genes between organisms; we are fundamentally limited by the historical adaptive

85   mutations and the epistatic relationships that define any particular organism. Before I discuss our success at creating such a

86   recombinant, I will address a question that has no doubt occurred to the reader: Why even bother performing such an

87   experiment with a reconstructed ancestral EF-Tu variant instead of a modern homolog from a different living organism?

4

88
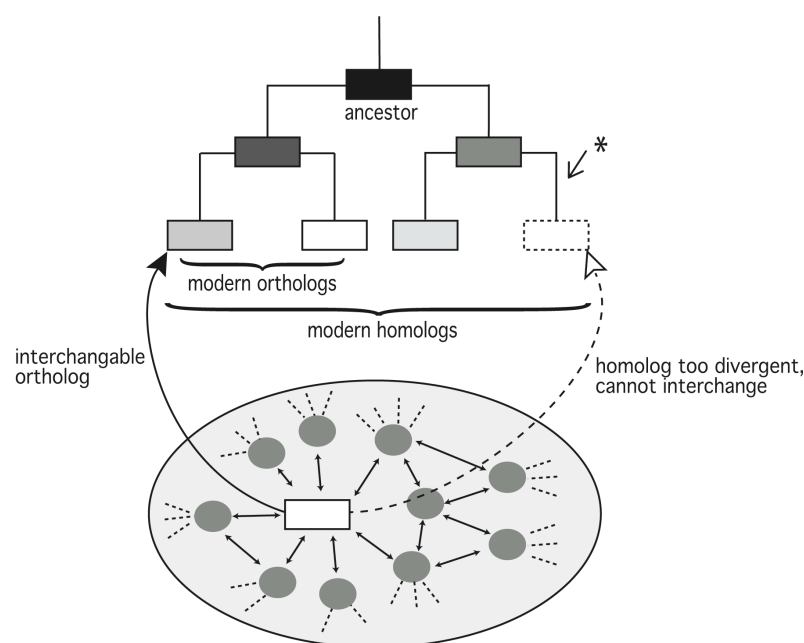
89   [**Figure 13.1**]

90   A simplified phylogenetic tree is shown. Phylogenetic trees help us to visualize the evolutionary relationships

91   among organisms. Modern gene and/or protein components are at the tip of the tree. Genes and/or proteins that are

92   connected by a branching point are more closely related to each other and may exhibit more closely related functional

93   properties. In ancestral sequence reconstruction, homologous characters of modern gene or protein components are used to

94   infer the character state of the common ancestor (located at the root of the tree, in black). A central component of a cell

95   (represented with the white box) will exhibit various interactions with other ancillary cellular partners (represented with

96   gray circles) that also have exhibit interactions with many other cellular components (represented with dashed lines). To

97   understand the evolutionary mechanisms that underlie protein function, we need to acknowledge that intrinsic and extrinsic

98   properties of the protein within its interaction network, not just the properties of an individual protein, contribute to and

99   perhaps define how a protein performs its function in a given environment. Replacing a modern gene with its modern

100   homologous or orthologous counterpart from another organism would allow us to identify the functional constraints that

101   shape the two proteins. However, such an approach cannot directly speak to how the replaced protein evolved because its

102   replacement is not on the same line of descent. By replacing a fine-tuned member of a networked system with another

103   component that has co-adapted to another networked system (shown in dashed box), the interactions between the replaced

104   protein and other native proteins may be damaged to the point of producing a non-functional system due to evolved

105   incompatibilities in the homolog (represented by an asterisk, *).]

106

5

107   **<A> Why Engineer a Modern Bacterial Genome with a Reconstructed Ancient Gene?**

108   One common way to assess how a protein may function in a foreign host is do so directly by removing that protein

109   from its host and inserting it into that foreign organism. One might ask why our new approach does not involve replacing a

110   modern gene with a homolog or ortholog from another modern organism. Ultimately, swapping a protein with its homolog

111   has been and will continue to provide valuable insights into functional constraints that shape the two proteins, regardless of

112   whether or not they share functional identity (Applebee et al. 2011; Couñago, Chen, and Shamoo 2006). However, this

113   approach has the drawback that swapped proteins may not share a direct line of descent that connects them over

114   evolutionary time (Figure 13.1). This lack of parity presents the possibility of "functional non-equivalence", meaning that

115   the two proteins may have traversed two separate and possibly functionally divergent adaptive paths that now prevent the

116   two homologs from functioning properly after being swapped between organisms (Figure 13.1). Moreover, it is important to

117   note that not all of the mutations of the modern homologs are adaptive; random mutations that are results of stochastic

118   events, for instance, can lead to the accrual of mutations in a modern homolog that could prevent the functionality of the

119   modern homolog in another organism once swapped, even while having been neutral under the conditions in which they

120   were accumulated (Camps et al. 2007; Romero and Arnold 2009).

121   Despite the problems involved, studies in which modern genes are replaced by resurrected ancestral genes are

122   worth the effort. In their article on the mechanistic approaches to study molecular evolution, Dean and Thornton remark

123   that "[t]he functional synthesis should move beyond studies of single genes to analyze the evolution of pathways and

124   networks that are made up of multiple genes. By studying the mechanistic history of the members of an interacting gene set,

125   it should be possible to reconstruct how metabolic and regulatory gene networks emerged and functionally diversified over

126   time" (Dean and Thornton 2007). It is not yet feasible to rewire a whole network but replacing a key component of a

127   complex network with its ancestral counterpart complements these efforts. It is expected that the ability of the interaction

128   network to function with the ancestral component will be limited by incompatibilities that have accrued by the rest of the

129   network over time since the point from which the ancestor was resurrected. A remaining question is how the modern

130   components of a cellular system would co-function with their ancestral interaction partners, if given a second chance.

131

132   **<A> An Ancient-Modern Hybrid Test Organism**

133   Following up on our *in vitro* work, we were set to generate a strain of modern *E. coli* harboring an approximately

134   500 million-year-old EF-Tu protein at the precise chromosomal location of the modern EF-Tu. The ancestral protein is

135   inferred to have been functioning within the common ancestor of γ-proteobacteria and has 21 out of 394 amino acid

6

136   differences with *E. coli* EF-Tu, and its melting temperature is comparable to *E. coli* EF-Tu (39.5 °C versus 37 °C,

137   respectively).

138   Marking the first genomic resurrection of a reconstructed ancestral essential gene in place of its modern

139   counterpart within an extant organism, we were able to obtain a viable *E. coli* strain that contains an ancient EF-Tu as the

140   sole genomic copy (Kacar and Gaucher 2012). We next measured the doubling time of the recombinant organism hosting

141   the ancestral EF-Tu. Consistent with our expectations, the hybrid organism exhibits lower fitness, as demonstrated by a

142   doubling time twice as long as its wild type parent strain (Figure 13.2). The hybrid is viable, showing that the ancient EF-Tu

143   can complement essential functions of a descendant in a modern organism. Moreover, the mutations that have accumulated

144   in the components of the modern bacteria over time do not prohibit the modern components that are essential for viability

145   from interacting with the ancient EF-Tu protein. This is an intriguing result, especially considering the vast number of

146   proteins EF-Tu interacts with beyond its primary interaction partners in the translation machinery. The hybrid is also less fit,

147   demonstrating that the ancestral component triggered a stress on the modern organism, creating an ideal case to monitor the

148   co-evolution of the ancient component and the recombinant bacteria.

149
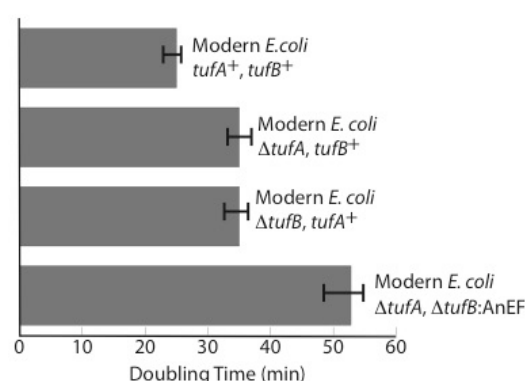
150   [**Figure 13.2**]

151   Replacement of a modern essential gene with its ancestral counterpart reduces host bacterium fitness. Precise

152   replacement of a modern bacterial EF-Tu gene with its ~500 million year old ancestor extends the bacterial strain REL606's

153   doubling time by two-fold. Two genes, *tufA* and *tufB*, (varying by just one amino acid) code for EF-Tu proteins in modern

154   *E. coli*. We deleted the *tufA* gene from the chromosome and replaced the *tufB* copy with the Ancient EF (referred in the

155   figure as ΔtufA, ΔtufB:AnEF). Deletions of *tufA* or *tufB* in the *E. coli* REL606 strain have similar effects (~ 34 minutes)

156   when deleted individually. Measurements are performed in rich growth media at 37°C in triplicates. Figure adapted from

157   (Kacar and Gaucher 2012).]

158

159   One interesting aspect of this modern/ancient recombinant approach is that it presents a new means of

160   understanding how protein function evolves as a consequence of changes in protein sequence, which in turn may allow us to

161   elucidate how changes in these functions correlate to the overall cellular context within living organisms. Although this is

162   not the primary motivation for setting up these experiments, the approach uniquely complements other studies focused on

163   understanding *in vivo* evolution of proteins. Here I would like to briefly discuss why this peripheral aspect carries biological

164   significance.

165   Modern cells are the products of immensely long evolutionary histories, and are thus the heirs of a biological

166   heritage that was shaped by the genetic and developmental characteristics of their ancestors, which were themselves shaped

167   by the environments in which they lived. Within this heritage are highly conserved proteins, the functions of which are so

168   crucial that the cellular machinery will not tolerate much change, and which therefore serve as functional fossils. Other

169   proteins are either highly resilient, of less crucial function, or else encode some other flexible potential, and so can readily

170   change in response to exigencies. These proteins show higher rates of evolution, and by definition, lower levels of

171   conservation over time and across taxa. The two groups may be readily distinguished by comparison of proteins across taxa

172   (Baker and Šali 2001; Kominek et al. 2013; Martí-Renom et al. 2000; Papp, Notebaart, and Pál 2011; Wellner, Raitses

173   Gurevich, and Tawfik 2013).

174   In order to assess how a protein performs its function, a basic approach is to couple molecular biology and

175   biochemistry by removing the protein from the cellular context, and measuring the protein's activity through its interactions

176   with a defined reactant in a test tube. To experimentally identify what role the *a priori* predicted functionally important sites

177   play, the predicted sites are altered via tools such as site-directed mutagenesis and the mutant protein's function is measured

178   *in vitro*. Two central points come from this; first, properly predicting the functionally important sites, second, properly

179   analyzing the protein's function so that studies reflect the protein's *in vivo,* and thus biologically-relevant, function in the

180   cell.

181   To address the first point, in order to predict what sites or domains of a protein carry functional importance, one

182   common way is to analyze the evolutionary conservancy of the protein sequence. This method briefly follows the alignment

183   of the amino acid sequences of multiple protein homologs, and then accessing the conserved, different and similar sites

184   through comparison of the aligned taxa. As expected, functionally important sites of a protein will be the most highly

185   conserved across homologs due to their crucial role in protein function (Benner 1989). However, this straightforward

186   methodology does not fully confront the fact that a protein site that is not necessarily conserved across homologous

187   organisms can also carry a functionally crucial role that goes beyond the immediate primary sequence of the protein itself,

188   and may represent an adaptation to a host's specific intracellular and extracellular environments (Fraser 2005; R. A. Jensen

189   1976; Khersonsky, Roodveldt, and Tawfik 2006). Recent approaches that include trajectory-scanning mutagenesis and

190   identify regions in which co-evolving proteins interact with each other provide valuable measures addressing this challenge

191   (Ashenberg and Laub 2013; Capra et al. 2010). Indeed, a single amino acid substitution far away from the primary

192   functional site can greatly impact that site's function (Copley 2003), indicative of various other factors shaping the

193   evolution of protein function.

194   This observations lead to the second point that needs to be carefully addressed when studying how proteins evolve.

195   In order to reveal how a protein performs its function by biologically realistic means, protein genotype and organismal

196   phenotype need to be directly connected. Engineering strains lacking a particular protein of interest and then reintroducing a

197   homolog or synthetically reconstructed variants of the protein in these mutant strains holds considerable value. Going one

198   step further and observing the co-adaptation between reconstructed proteins and microbial organisms through laboratory

199   evolution would allow us to examine the protein's function *in vivo*. Finally, this approach would allow us to identify

200   functionally important sites that are not necessarily conserved across taxa but are specific to a host's intercellular

201   environment, and thus allowing us to consider the context-dependent protein adaptation within a specific lineage.

202

203   **<A> Experimental Evolution of Bacteria Containing an Ancient Gene Component**

204   The above discussion provides a context for understanding how functionality is connected to protein divergence.

205   Only after site-specific evolutionary constraints are defined, can we only truly begin to understand how a protein

206   'functions'. One alternative way to determine protein functionality is to let an organism mutate its proteome in response to

207   some intra- or intercellular environmental pressure. Such pressures can arise from the availability of novel energy sources or

208   even from the manipulation of internal cellular components. Towards this end, experimental evolution with microbes would

209   provide information on the organismal level by monitoring the real-time evolution of microbial populations to some unique

210   pressure. In these evolution experiments, microbial populations are evolved in the lab through either continuous culture or

211   serial transfer under controlled conditions, permitting evolution to be studied in unprecedented detail in near real time

212   (Elena and Lenski 2003).

213   The experimental evolution approach is particularly powerful because of the high level of control it permits.

214   Samples of evolving populations may be frozen at regular intervals while still remaining viable, providing "frozen fossil

215   records" that be analyzed to examine a variety of important questions (Elena and Lenski 2003). Experimental evolution may

216   be used for highly detailed study of a variety of questions, such as whether evolution follows contingent or deterministic

217   paths, whether mutations accumulate neutrally or adaptively, and how epistatic interactions impact evolution. Indeed,

218   laboratory evolution experiments using microbes have provided deep insights into the interplay between contingency and

219   deterministic processes in evolution (Blount, Borland, and Lenski 2008; Blount this volume; Losos 2011; Travisano et al.

220   1995; Vermeij 2006).

221   The experimental evolution approach is key in our next step with the ancient-modern hybrid organism we have

222   constructed to replay the tape. We are currently evolving the reconstructed organism in the laboratory to study how

223   compensatory evolution alters the ancient gene and members of its epistatic network. Of particular interest is how much and

224   how quickly the construct's fitness and phenotype will change, and how much of these changes are mediated by the ancient

225   EF-Tu's direct accumulation of beneficial mutations versus their accumulation in other, associated genes. Moreover, we

226   wish to assess the extent to which the mutational trajectory is guided by random and unpredictable events (i.e., attributable

227   to chance) and how much it is determined directly by the genotype and the phenotype of the ancient EF-Tu engineered

228   inside the bacteria (i.e., attributable to contingency).

229   Evolving a bacterium containing a single ancient gene in its genome is analogous to replaying a particular track on

230   the tape of life within the context of the modern organism. To play with the metaphor a bit, we made a mix-tape by splicing

231   a very old track from the tape of life, and are now playing it on a modern tape deck (the modern organism). Both the tape

232   and the tape deck can evolve in this case, and we will be able to examine how they evolve and adapt to each other,

233   potentially recapitulating some of the evolution that occurred in the first place. While this sort of replay experiment is not

234   exactly what Gould had in mind, how the organism adapts to the ancestral protein and vice versa promises to shed light on

235   the molecular evolution that took place during EF-Tu evolution.

236   This approach does have limitations as well. It is important to recognize that the ancient component resurrected in

237   a modern organism's adaptation will be shaped by its interactions with modern components adapted to both its modern

238   counterpart, and a cellular environment that is millions of years from ahead of the conditions of the ancient component. It is

239   also important to note that the environmental conditions in the laboratory are unlikely to be comparable to the conditions

240   and circumstances under which the ancestral protein evolved into its modern descendant.

241   Despite its limitations, re-evolving ancient genes in modern organisms permits us to investigate historical aspects

242   of repeatability and parallelism in evolution that replacing a gene with a modern homolog does not. Multiple aspects of the

243   evolutionary processes, including mutation and genetic drift, are inherently stochastic, which makes it challenging to predict

244   how a system hosting a reconstructed gene (as well as a modern homolog) would co-evolve. By setting up laboratory

245   evolution experiments of microbial organisms that carry an ancient component in replicate populations, we are able to

246 observe whether or not multiple populations carrying the ancient component evolve the same way, or in parallel ways,

247 under identical environmental conditions. Engineering constructs with genes encoding different reconstructed ancestral

248 states of the same protein (e.g., 500 million-year-old, 1 billion-year-old, etc.) would permit investigation of whether or not

249 evolutionary trajectories are strictly dependent on evolutionary starting points (Figure 13.3; see also Travisano et al. 1995).

250 For instance, certain adaptive zones may readily be more accessible as we go further back in time; i.e., the older the

251 component we resurrect, the greater the likelihood of finding a mutational point that is capable of functional innovation by

252 resetting the epistatic ratchet (Bridgham, Ortlund, and Thornton 2009).

253

254
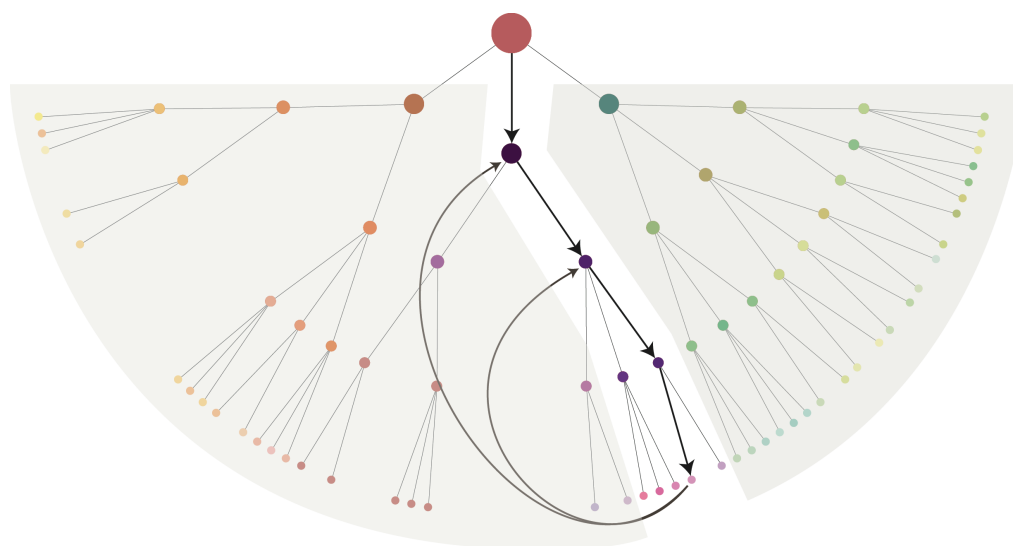
255 [**Figure 13.3**]:

256 A hypothetical scheme of evolutionary patterns: Interaction between an organism containing ancestral variants of a

257 protein (represented by different colored circles) and its environment will produce a series of adaptive zones. Replacing a

258 modern protein with an ancestral counterpart may potentially lead the organism to explore a wider potential adaptive

259 landscape. Replaying evolution starting from successively older time-points (shown with successively longer arrows) would

260 then allow access to points corresponding in time to historic evolutionary states that preceded the accumulation of

261 mutational steps, thus allowing the engineered organism to operate in different adaptive zones.]

262

263

264

265

266

11

267 **<A> Concluding Remarks**

268   To directly examine the relationship between historical constraints and evolutionary trajectories, and to assess the

269 role historical contingency played in shaping protein evolution within a biologically realistic framework, it is important to

270 determine the historical paths of proteins, both genotypic and phenotypic levels, and then study how these paths affected the

271 organism. Many properties of protein-protein interaction networks, not just those of individual proteins, contribute to and

272 perhaps define how a protein performs its function in a given environment. Consequently, the biochemistry of proteins may

273 be intimately linked to their host organisms' behavior.

274   The only lesson history teaches is that no one learns from history (Hegel 1953). This playful quote is not the case

275 for understanding the role history plays in shaping evolution. Integrating molecular-level theory into studies of evolutionary

276 history and merging ancestral sequence reconstruction studies with experimental evolution will help to provide mechanistic

277 explanations into several long-standing questions in evolutionary biology: How does an organism's history shape its future

278 trajectories, and are there deterministic paths along these trajectories? Does evolution inevitably lead to the same endpoints?

279 How much does the past restrict the future? Moreover, how do answers to these questions depend on the genetic and the

280 environmental conditions of the system examined? Synthesis of the fields of synthetic biology, biochemistry, and

281 experimental evolution holds great promise for generating a new understanding of functional, structural and historical

282 constraints that shape biological evolution. Indeed, this synthesis holds great promise for learning much of and from life's

283 history.

284

285   **<A> Acknowledgements**

292

293

294

295

12

296   **<A> References**

Applebee, M. Kenyon, Andrew R. Joyce, Tom M. Conrad, Donald W. Pettigrew, and Bernhard Ø Palsson. 2011. "Functional and Metabolic Effects of Adaptive Glycerol Kinase (GLPK) Mutants in *Escherichia Coli*." *Journal of Biological Chemistry* 286 (26): 23150–59. doi:10.1074/jbc.M110.195305.

Ashenberg, Orr, and Michael T. Laub. 2013. "Using Analyses of Amino Acid Coevolution to Understand Protein Structure and Function." In *Methods in Enzymology*, edited by Amy E. Keating, 523:191–212. Methods in Protein Design. London: Academic Press.

Baker, David, and Andrej Šali. 2001. "Protein Structure Prediction and Structural Genomics." *Science* 294 (5540): 93–96. doi:10.1126/science.1065659.

Benner, Steven A. 1989. "Patterns of Divergence in Homologous Proteins as Indicators of Tertiary and Quaternary Structure." *Advances in Enzyme Regulation* 28: 219–36. doi:10.1016/0065-2571(89)90073-3.

Benner, Steven A., Slim O. Sassi, and Eric A. Gaucher. 2007. "Molecular Paleoscience: Systems Biology from the Past." In *Advances in Enzymology*, edited by Eric J. Toone, 1–132. Hoboken, NJ: Wiley.

Blount, Zachary D., Christina Z. Borland, and Richard E. Lenski. 2008. "Historical Contingency and the Evolution of a Key Innovation in an Experimental Population of *Escherichia Coli*." *Proceedings of the National Academy of Sciences* 105 (23): 7899–7906. doi:10.1073/pnas.0803151105.

Bridgham, Jamie T., Eric A. Ortlund, and Joseph W. Thornton. 2009. "An Epistatic Ratchet Constrains the Direction of Glucocorticoid Receptor Evolution." *Nature* 461 (7263): 515–19. doi:10.1038/nature08249.

Camps, Manel, Asael Herman, Ern Loh, and Lawrence A. Loeb. 2007. "Genetic Constraints on Protein Evolution." *Critical Reviews in Biochemistry and Molecular Biology* 42 (5): 313–26. doi:10.1080/10409230701597642.

Capra, Emily J., Barrett S. Perchuk, Emma A. Lubin, Orr Ashenberg, Jeffrey M. Skerker, and Michael T. Laub. 2010. "Systematic Dissection and Trajectory-Scanning Mutagenesis of the Molecular Interface That Ensures Specificity of Two-Component Signaling Pathways." *PLoS Genetics* 6 (11): e1001220. doi:10.1371/journal.pgen.1001220.

Cole, Megan F., and Eric A. Gaucher. 2011. "Exploiting Models of Molecular Evolution to Efficiently Direct Protein Engineering." *Journal of Molecular Evolution* 72 (2): 193–203. doi:10.1007/s00239-010-9415-2.

Copley, Shelley D. 2003. "Enzymes with Extra Talents: Moonlighting Functions and Catalytic Promiscuity." *Current Opinion in Chemical Biology* 7 (2): 265–72. doi:10.1016/S1367-5931(03)00032-2.

Couñago, Rafael, Stephen Chen, and Yousif Shamoo. 2006. "In Vivo Molecular Evolution Reveals Biophysical Origins of Organismal Fitness." *Molecular Cell* 22 (4): 441–49. doi:10.1016/j.molcel.2006.04.012.

Czworkowski, John, and Peter B. Moore. 1996. "The Elongation Phase of Protein Synthesis." *Progress in Nucleic Acid Research and Molecular Biology* 54: 293–332.

Dean, Antony M., and Joseph W. Thornton. 2007. "Mechanistic Approaches to the Study of Evolution: The Functional Synthesis." *Nature Reviews Genetics* 8 (9): 675–88. doi:10.1038/nrg2160.

Defeu Soufo, Hervé Joël, Christian Reimold, Uwe Linne, Tobias Knust, Johannes Gescher, and Peter L. Graumann. 2010. "Bacterial Translation Elongation Factor EF-Tu Interacts and Colocalizes with Actin-like MreB Protein." *Proceedings of the National Academy of Sciences* 107 (7): 3163–68. doi:10.1073/pnas.0911979107.

Elena, Santiago F., and Richard E. Lenski. 2003. "Evolution Experiments with Microorganisms: The Dynamics and Genetic Bases of Adaptation." *Nature Reviews Genetics* 4 (6): 457–69. doi:10.1038/nrg1088.

Fortuna, Miguel A., Luis Zaman, Aaron P. Wagner, and Charles Ofria. 2013. "Evolving Digital Ecological Networks." *PLoS Computational Biology* 9 (3): e1002928. doi:10.1371/journal.pcbi.1002928.

Fraser, Hunter B. 2005. "Modularity and Evolutionary Constraint on Proteins." *Nature Genetics* 37 (4): 351–52. doi:10.1038/ng1530.

Galtier, Nicolas, Nicolas Tourasse, and Manolo Gouy. 1999. "A Nonhyperthermophilic Common Ancestor to Extant Life Forms." *Science* 283 (5399): 220–21. doi:10.1126/science.283.5399.220.

Gaucher, Eric A. 2007. "Ancestral Sequence Reconstruction as a Tool to Understand Natural History and Guide Synthetic Biology: Realizing and Extending the Vision of Zuckerkandl and Pauling." In *Ancestral Sequence Reconstruction*, edited by David A. Liberles, 20–33. Oxford: Oxford University Press.

Gaucher, Eric A., Sridhar Govindarajan, and Omjoy K. Ganesh. 2008. "Palaeotemperature Trend for Precambrian Life Inferred from Resurrected Proteins." *Nature* 451 (7179): 704–7. doi:10.1038/nature06510.

Gould, Stephen Jay. 1989. *Wonderful Life*. New York and London: W. W. Norton & Company.

Gromiha, M. Michael, Motohisa Oobatake, and Akinori Sarai. 1999. "Important Amino Acid Properties for Enhanced Thermostability from Mesophilic to Thermophilic Proteins." *Biophysical Chemistry* 82 (1): 51–67. doi:10.1016/S0301-4622(99)00103-9.

Harms, Michael J., and Joseph W. Thornton. 2013. "Evolutionary Biochemistry: Revealing the Historical and Physical Causes of Protein Properties." *Nature Reviews Genetics* 14 (8): 559–71. doi:10.1038/nrg3540.

Hegel, Georg Wilhelm Friedrich. 1953. *Reason in History*. Translated by Robert S. Hartman. New York: Liberal Arts Press.

Jensen, Roy A. 1976. "Enzyme Recruitment in Evolution of New Function." *Annual Review of Microbiology* 30: 409–25. doi:10.1146/annurev.mi.30.100176.002205.

Kacar, Betul, and Eric A. Gaucher. 2012. "Towards the Recapitulation of Ancient History in the Laboratory: Combining Synthetic Biology with Experimental Evolution." In *Artificial Life 13*, edited by Christoph Adami, David M. Bryson, Charles Ofria, and Robert T. Pennock, 11–18. Cambridge, MA: The MIT Press.

Kacar, Betul, and Eric A. Gaucher. 2013. "Experimental Evolution of Protein-Protein Interaction Networks." *Biochemical Journal* 453 (3): 311–19. doi:10.1042/BJ20130205.

Khersonsky, Olga, Cintia Roodveldt, and Dan S. Tawfik. 2006. "Enzyme Promiscuity: Evolutionary and Mechanistic Aspects." *Current Opinion in Chemical Biology* 10 (5): 498–508. doi:10.1016/j.cbpa.2006.08.011.

Kominek, Jacek, Jaroslaw Marszalek, Cécile Neuvéglise, Elizabeth A. Craig, and Barry L. Williams. 2013. "The Complex Evolutionary Dynamics of Hsp70s: A Genomic and Functional Perspective." *Genome Biology and Evolution* 5 (12): 2460–77. doi:10.1093/gbe/evt192.

Kratzer, James T., Miguel A. Lanaspa, Michael N. Murphy, Christina Cicerchi, Christina L. Graves, Peter A. Tipton, Eric A. Ortlund, Richard J. Johnson, and Eric A. Gaucher. 2014. "Evolutionary History and Metabolic Insights of Ancient Mammalian Uricases." *Proceedings of the National Academy of Sciences*, 201320393. doi:10.1073/pnas.1320393111.

Losos, Jonathan B. 1994. "Historical Contingency and Lizard Community Ecology." In *Lizard Ecology: Historical and Experimental Perspectives*, edited by Laurie J. Vitt and Eric R. Planka, 319–33. Princeton, NJ: Princeton University Press.

———. 2011. "Convergence, Adaptation, and Constraint." *Evolution* 65 (7): 1827–40. doi:10.1111/j.1558-5646.2011.01289.x.

Martí-Renom, Marc A., Ashley C. Stuart, András Fiser, Roberto Sánchez, Francisco Melo, and Andrej Šali. 2000. "Comparative Protein Structure Modeling of Genes and Genomes." *Annual Review of Biophysics and Biomolecular Structure* 29: 291–325. doi:10.1146/annurev.biophys.29.1.291.

Ogawa, Tomohisa, and Tsuyoshi Shirai. 2013. "Experimental Molecular Archeology: Reconstruction of Ancestral Mutants and Evolutionary History of Proteins as a New Approach in Protein Engineering." In *Protein Engineering - Technology and Application*, edited by Tomohisa Ogawa. InTech.

Papp, Balázs, Richard A. Notebaart, and Csaba Pál. 2011. "Systems-Biology Approaches for Predicting Genomic Evolution." *Nature Reviews Genetics* 12 (9): 591–602. doi:10.1038/nrg3033.

Pauling, Linus, and Emile Zuckerkandl. 1963. "Chemical Paleogenetics: Molecular 'Restoration Studies' of Extinct Forms of Life." *Acta Chemica Scandinavia* 17 (Suppl 1): S9–16.

Pieper, Rembert, Quanshun Zhang, David J. Clark, Shih-Ting Huang, Moo-Jin Suh, John C. Braisted, Samuel H. Payne, Robert D. Fleischmann, Scott N. Peterson, and Saul Tzipori. 2011. "Characterizing the *Escherichia Coli* O157:H7 Proteome Including Protein Associations with Higher Order Assemblies." *PLoS ONE* 6 (11): e26554. doi:10.1371/journal.pone.0026554.

Risso, Valeria A., Jose A. Gavira, Diego F. Mejia-Carmona, Eric A. Gaucher, and Jose M. Sanchez-Ruiz. 2013. "Hyperstability and Substrate Promiscuity in Laboratory Resurrections of Precambrian β-Lactamases." *Journal of the American Chemical Society* 135 (8): 2899–2902. doi:10.1021/ja311630a.

Romero, Philip A., and Frances H. Arnold. 2009. "Exploring Protein Fitness Landscapes by Directed Evolution." *Nature Reviews Molecular Cell Biology* 10 (12): 866–76. doi:10.1038/nrm2805.

Shimizu, Yoshihiro, Akio Inoue, Yukihide Tomari, Tsutomu Suzuki, Takashi Yokogawa, Kazuya Nishikawa, and Takuya Ueda. 2001. "Cell-Free Translation Reconstituted with Purified Components." *Nature Biotechnology* 19 (8): 751–55. doi:10.1038/90802.

Stackhouse, Joseph, Scott R. Presnell, Gerard M. McGeehan, Krishnan P. Nambiar, and Steven A. Benner. 1990. "The Ribonuclease from an Extinct Bovid Ruminant." *FEBS Letters* 262 (1): 104–6. doi:10.1016/0014-5793(90)80164-E.

Travisano, Michael, Judith A. Mongold, Albert F. Bennett, and Richard E. Lenski. 1995. "Experimental Tests of the Roles of Adaptation, Chance, and History in Evolution." *Science* 267 (5194): 87–90. doi:10.1126/science.7809610.

Vermeij, Geerat J. 2006. "Historical Contingency and the Purported Uniqueness of Evolutionary Innovations." *Proceedings of the National Academy of Sciences* 103 (6): 1804–9. doi:10.1073/pnas.0508724103.

Wellner, Alon, Maria Raitses Gurevich, and Dan S. Tawfik. 2013. "Mechanisms of Protein Sequence Divergence and Incompatibility." *PLoS Genetics* 9 (7): e1003665. doi:10.1371/journal.pgen.1003665.

Zhou, Ying, Haruichi Asahara, Eric A. Gaucher, and Shaorong Chong. 2012. "Reconstitution of Translation from *Thermus Thermophilus* Reveals a Minimal Set of Components Sufficient for Protein Synthesis at High Temperatures and Functional Conservation of Modern and Ancient Translation Components." *Nucleic Acids Research* 40 (16): 7932–45. doi:10.1093/nar/gks568.

297