Effects of metabolic rate on protein evolution

James F. Gillooly1, 2, Michael W. McCoy1 and Andrew P. Allen2

1Department of Zoology, University of Florida, Gainesville, FL 32611, USA
2National Center for Ecological Analysis and Synthesis, Santa Barbara, CA 93101, USA

1Author for correspondence (gillooly@zoology.ufl.edu).

Since the modern evolutionary synthesis was first proposed early in the twentieth century, attention has focused on assessing the relative contribution of mutation versus natural selection on protein evolution. Here we test a model that yields general quantitative predictions on rates of protein evolution by combining principles of individual energetics with Kimura's neutral theory. The model successfully predicts much of the heterogeneity in rates of protein evolution for diverse eukaryotes (i.e. fishes, amphibians, reptiles, birds, mammals) from different thermal environments. Data also show that the ratio of non-synonymous to synonymous nucleotide substitution is independent of body size, and thus presumably of effective population size. These findings indicate that rates of protein evolution are largely controlled by mutation rates, which in turn are strongly influenced by individual metabolic rate.

Keywords: metabolic theory; neutral theory; mutation; scaling; molecular evolution

1. INTRODUCTION

Variation in mutation rates is often invoked to explain heterogeneity in rates of non-coding DNA evolution, as predicted by Kimura’s neutral theory (Kimura 1983; Kumar 2005). While some research has pointed to a link between mutation and protein evolution (e.g. Stanley & Harrison 1999), other studies indicate that protein evolution is controlled primarily by natural selection (Gillespie 1991; McDonald & Kreitman 1991; Sawyer 2003). Resolution of this issue has been hampered by our limited understanding of the factors that control mutation rates (Bromham & Penny 2003).

Recent studies have shed light on this issue by showing that mutation rate is strongly influenced by metabolic rate and the factors that control it (Martin & Palumbi 1993; Allen et al. 2006; Wright et al. 2006). In particular, Gillooly et al. (2005) presented evidence that variation in synonymous rates of nucleotide substitution can be predicted based on the body size and temperature dependence of metabolic rate. However, it is unclear whether interspecific differences in metabolic rate, due to differences in size and temperature, can also explain variation in rates of protein evolution. Here we examine this question by first extending the model of Gillooly et al. (2005) to yield predictions on rates of protein evolution, and then testing these predictions using new data compiled for diverse species from different environmental temperatures.

2. BACKGROUND

The Gillooly et al. (2005) model incorporates the effects of body size and temperature on mass-specific metabolic rate, \( \dot{B} (W g^{-1}) \), using the following equation (Gillooly et al. 2001):

\[
\dot{B} = b_0 M^{-1/4} e^{-E / k T},
\]

(2.1)

where \( M \) is size (g), \( T \) is temperature in Kelvin (K); and \( b_0 \) is a normalization constant independent of size and temperature (W g\(^{-3/4}\)). The 1/4-power scaling of metabolic rate with size, and the exponential increase in metabolic rate with temperature, described by the Boltzmann–Arrhenius factor, \( e^{-E / k T} \), are well supported by both theory and data (Peters 1983; West et al. 1997; Gillooly et al. 2001). Here \( E \) is the activation energy of the respiratory complex (approx. 0.6–0.7 eV) and \( k \) is Boltzmann’s constant (8.62 \times 10^{-5} eV K\(^{-1}\)).

Gillooly et al. (2005) formulated their model of molecular evolution using equation (2.1) by invoking two simplifying assumptions. First, point mutations occur at a rate proportional to mass-specific metabolic rate. This assumption is consistent with the hypotheses that mutation rates are governed by the production of oxidative free radicals, and the hypothesis that mutation rates are governed by errors that occur during germline replication (see Martin & Palumbi 1993). Second, the point mutations fixed in populations are selectively neutral, which is reasonable, provided that the selection coefficient is less than half the reciprocal of the effective population size for most mutations (Kimura 1983). Given these assumptions, their model predicts that the synonymous substitution rate per unit absolute time, \( \alpha_s \) (substitutions per nucleotide Myr\(^{-1}\)), should be controlled by metabolic rate, \( \dot{B} \), through its effects on the point mutation rate, \( v \dot{B} \) (mutations per nucleotide - Myr\(^{-1}\)), as follows:

\[
\alpha_s = f_v v \dot{B} = f_v v b_0 M^{-1/4} e^{-E / k T},
\]

(2.2)

where \( v \) is the number of mutations per nucleotide per unit energy flux through a gram of tissue (mutations per nucleotide g J\(^{-1}\)) and \( f_v \) is the fraction of synonymous mutations that are neutral. Gillooly et al. (2005) obtained support for equation (2.2) using extensive nuclear and mitochondrial data (see also Estabrook et al. 2007). As explained by Gillooly et al. (2005), since generation times have the same size and temperature dependence as metabolic rate, their model predicts that the mutation rate per generation is independent of size and temperature.

3. EXTENDING THE GILLOOLY ET AL. (2005) MODEL TO PROTEIN EVOLUTION

Equation (2.2) can be extended to encompass protein evolution by assuming that most changes to amino acid sequences are also the result of neutral mutations. This assumption implies that the non-synonymous rate of nucleotide substitution, \( \alpha_{ns} \), should be proportional to the synonymous rate, \( \alpha_s \).
and should therefore show the same body size and temperature dependence as metabolic rate, \( B \),

\[
\alpha_n = (f_n/f_s)\alpha_a = f_n v b_0 M^{-1/4} e^{-E/kT},
\]

(3.1)

where \( f_n \) is the fraction of non-synonymous mutations that are selectively neutral. Equation (3.1) implies that protein evolution, \( \alpha_a \) (amino acid substitutions per amino acid Myr\(^{-1}\)), should also be governed by metabolic rate,

\[
\alpha_a = f_n v_a B = f_n v_b b_0 M^{-1/4} e^{-E/kT},
\]

(3.2)

where \( f_a \) is the fraction of amino acid changes that are neutral and \( v_a \) is the number of amino acid changes per unit energy flux. Given that both \( v \) and \( v_a \) are independent of size and temperature, equations (3.1) and (3.2) predict that rates of protein evolution, like synonymous rates of DNA substitution, ‘tick’ at a constant rate per unit energy rather than per unit time.

Equations (3.1) and (3.2) yield two predictions on rates of protein evolution. First, for a given gene or protein, equations (3.1) and (3.2) predict that the logarithms of size-corrected rates of non-synonymous nucleotide and amino acid substitution, \( \ln(\alpha M^{1/4}) \), should be linear functions of inverse absolute temperature, \( 1/kT \), with slopes of \(-E \approx -0.65 \text{ eV}\).

Second, logarithms of temperature-corrected rates, \( \ln(\alpha e^{E/kT}) \), should be linear functions of the logarithm of body size, \( \ln(M) \), with slopes of \(-1/4\).

We evaluated these predictions using protein evolution data compiled for a diverse suite of endotherms and ectotherms that span a range of body sizes (approx. \( 1 \text{–} 10^7 \text{ g} \)) and temperatures...
We performed analyses for two mitochondrial genes (nicotinamide adenine dinucleotide, i.e. NADH, and cytochrome b) and one family of nuclear genes (haemoglobin). To the best of our knowledge, these were the only protein-coding genes with sufficient data to perform these analyses. Evolutionary rates were calculated as $\alpha = D/2G$ by combining estimated genetic distances between taxon pairs, $D$, with divergence times, $G$, that were independently estimated from fossil records or well-established biogeographic events (electronic supplementary materials 2–4). Body sizes and temperatures were estimated using methods described by Gillooly et al. (2005) and Allen et al. (2006; see electronic supplementary material 2). Unfortunately, data on the field metabolic rates of the species considered here were not available.

The data are largely supportive of the model predictions. Multiple regression analyses indicate that both body size and temperature have significant independent effects on rates of protein evolution for the mitochondrial genes NADH and cytochrome b (temperature, $p < 0.04$; body size, $p < 0.001$), even

Figure 2. Predicted linear relationship between rates of amino acid and synonymous DNA evolution for (a) NADH, (b) cytochrome b and (c) haemoglobin. Fitted lines are forced through the origin using model-I regression. Ratio of non-synonymous to synonymous substitution versus the logarithm of body mass for (d) NADH, (e) cytochrome b and (f) haemoglobin. As predicted, the fitted slopes are not significantly different from 0 in all cases ($p > 0.08$).
though these variables are correlated for both datasets (r=0.55 and 0.45, respectively; p<0.05). Moreover, body size-corrected rates of protein evolution are linearly related to inverse absolute temperature for both mitochondrial genes (figure 1a,b), with 95% confidence intervals (CI) for model-I and model-II regression slopes that include the predicted value of $-E_{a} = -0.65$ eV (electronic supplementary material 1). Furthermore, temperature-corrected rates of amino acid evolution show significant linear relationships with the logarithm of body size for both of the mitochondrial genes and the haemoglobin nuclear gene family (figure 1c–e). The model-I slopes of these relationships have 95% CI that include the predicted value of $-0.25$ for haemoglobin ($-0.25$ to $-0.07$), but are slightly shallower for NADH ($-0.24$ to $-0.11$) and cytochrome b ($-0.22$ to $-0.06$). However, the model-II slopes have 95% CI that include $-0.25$ in all cases (electronic supplementary material 1). Since the model-I slopes (but not model-II slopes) are consistently less than $-0.25$, and larger organisms generally have smaller effective population sizes (Lynch & Conery 2003), we cannot rule out the possible influence of effective population size, as predicted by the ‘nearly neutral’ theory (Ohta 1973). The predicted relationships (dashed lines in figure 1) explain up to 53% of observed variation in rates of protein evolution, indicating that body size and temperature impose significant constraints (electronic supplementary material 1).

Combining equations (2.1), (3.1) and (3.2) yields two additional predictions. First, the rate of protein evolution, $a_{n}/a_{k}$ should increase linearly with the synonymous substitution rate, $a_{n}$, with a slope of $(f_{s}/f_{a})(a_{n}/v)$ and an intercept of 0. This prediction applies across species, irrespective of size or temperature. Second, the ratio of non-synonymous to synonymous substitution rates, $a_{d}/a_{s}$ (i.e. $d_{v}/d_{o}$), which is often thought to characterize the intensity of natural selection (Nei & Gojobori 1986), is predicted to be independent of body size and temperature such that $a_{d}/a_{s} = f_{d}/f_{a}$.

The data support these two additional predictions. First, the rate of amino acid substitution increases linearly with the rate of synonymous DNA substitution for NADH, cytochrome b and haemoglobin (figure 2a–c). The synonymous rate accounts for up to 94% of the heterogeneity in rates of protein evolution among species, consistent with previous research for individual taxonomic groups (Stanley & Harrison 1999). Second, for all genes, the ratio of non-synonymous to synonymous substitution is independent of body size over seven orders of magnitude (figure 2d–f). This result is surprising because effective population size declines with increasing body size (Lynch & Conery 2003) and is expected to influence evolutionary dynamics for genes under positive selection (Ohta 1973; Lynch & Conery 2003; Keightley et al. 2005). Thus, this result suggests that evolutionary rates of the protein-coding genes considered here are largely independent of effective population size, and thus are controlled mainly by neutral processes.

We are not implying that body size and temperature are the only factors that influence rates of protein evolution. Other contributing factors may include base pair composition, transposition and the efficiency of DNA repair (Pal et al. 2006). Some combination of these and other factors may help explain the substantial residual variation in figures 1 and 2. Differences in the intercepts may also reflect variation in the fraction of mutations that are neutral, $f_{n}$ and variation in mutation rates within and among nuclear and mitochondrial genomes (Li 1997). Moreover, mutation rate may not play a significant role in protein evolution for genes under strong selective pressure. Recent work suggests that while some species may have little or no positive selection operating, other species may have a large fraction of loci under some form of positive selection (Eyre-Walker 2006).

Still, the results presented here support the hypothesis that rates of protein evolution are largely controlled by mutation rates, which in turn are strongly influenced by individual metabolic rate. Overall, these findings indicate that rates of molecular evolution will vary predictably among natural communities that comprise species that differ in body size and/or temperature (see Allen et al. 2006). However, the extent to which amino acid changes lead to phenotypic change remains unclear. Some recent work indicates that such changes often alter the structure and function of proteins, and thereby lead to phenotypic evolution (Bloom et al. 2005; Ma et al. 2006). While such changes may have no direct effect on fitness over the short term, they may nevertheless provide the raw materials for future innovation. Thus, body size and temperature, through their effects on individual metabolism, may play an important role in both genotypic and phenotypic evolution.

We thank C. Baer, E. Braun, K. McCoy, R. Kimball, J. Allen and M. Miyamoto for their helpful discussions. A.P.A. was supported by NCEAS, which is funded by NSF grant DEB-0072909, and UC Santa Barbara.


NOTICE OF CORRECTION

Figures 1 and 2 are now correct.  

8 October 2007