

Codon bias coevolves with longevity

Krisztina Kerekes

HUN-REN Research Centre for Natural Sciences

Mária Trexler

HUN-REN Research Centre for Natural Sciences

László Bányai

HUN-REN Research Centre for Natural Sciences

László Patthy

`patthy.laszlo@ttk.hu`

HUN-REN Research Centre for Natural Sciences

Article

Keywords: aging, longevity, multidomain proteins, protein folding, proteostasis, translation fidelity

Posted Date: July 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9835155/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Somatic mutations drive carcinogenesis and aging, shortening animal lifespan. Since gene vulnerability strongly depends on the abundance of mutation hotspots, we tested whether negative selection of hypermutable (e.g., CpG-bearing) codons could contribute to the evolution of mammalian longevity. Our studies showed that the CGA codon was significantly more depleted in long-lived than short-lived mammals, suggesting negative selection of this hypermutable stopogenic codon. Interestingly, our analyses revealed lifespan-dependent codon-usage changes in most amino acids. For a few amino acids, altered codon usage favored translationally optimal codons in long-lived animals, reducing mistranslation and abnormal protein formation. Surprisingly, for a larger group of amino acids, codon usage in long-lived animals favored translationally nonoptimal codons lacking matching isodecoder tRNAs. The most likely explanation for this observation is that slower translation at these codons facilitates co-translational folding, thereby reducing the chances of misfolding and aggregation of misfolded proteins in long-lived animals. Our results suggest that changes in codon usage may produce more balanced proteostasis and a lower rate of cellular aging in long-lived animals. Our finding is consistent with the notion that a hallmark of aging is loss of proteostasis, manifested in the accumulation of abnormal, misfolded proteins.

Significance

Recent studies have shown that rodent species with longer lifespan have increased fidelity of protein synthesis, leading to the hypothesis that increased accuracy of translation and lower rate of formation of abnormal proteins plays a key role in the evolution of longevity. Although this hypothesis implies that evolution of longevity may be driven by changes in codon usage favoring translationally optimal codons, our studies on rodents and other mammals have revealed that in the case of the majority of amino acids translationally non-optimal codons are enriched in long-lived species. We suggest that the use of non-optimal codons may have significant benefits: slowdown of translation at these codons favors correct co-translational folding by allowing the nascent chain more time to fold and the reduced rate of misfolding and protein aggregation result in lower rate of cellular aging in long-lived species.

1. Introduction

Somatic mutations accumulate in protein-coding genes throughout life: they are known to drive carcinogenesis [1] and have long been suspected to play a key role in aging [2], [3].

The chances of acquiring nonsense, missense and silent mutations by a protein-coding gene are proportional, not only to its length, but also to the number of hypermutable sites, such as CpG dinucleotides. In our earlier work we have shown that the hypermutability of CpG bearing CGN codons of arginine contribute significantly to the vulnerability of cancer genes [4], [5]. The CGA codon of arginine is particularly dangerous since, due to hypermutability of its CpG dinucleotide, it generates a nonsense codon (TGA) at an exceptionally high rate [4]. In the present work we have tested whether negative

selection of hypermutable CpG bearing codons could play a role in the evolution of longevity in mammalian taxa. We have monitored possible signals of selection of synonymous codons by analyzing correlations of codon bias with the lifespan of mammalian species.

Our studies have shown that only the hypermutable CpG bearing stopogenic Arg codon, CGA, was significantly more depleted in long-lived than short-lived mammals, suggesting negative selection of this codon during evolution of longevity.

Unexpectedly, our analyses have revealed remarkable correlations between evolution of longevity in mammals and changes in usage of synonymous codons of amino acids. In the case of a few amino acids the change in codon usage favoured translationally optimal codons, i.e. those that match the most abundant isodecoder tRNAs, encoded by the highest number of tRNA genes in mammals. The most likely explanation for this observation is that these translationally optimal codons ensure higher accuracy (lower rate of mistranslation), thereby reducing the chances of the formation of abnormal proteins in long-lived animals [6], [7]. This explanation is consistent with the fact that one of the most important hallmarks of aging is loss of proteostasis, manifested in the accumulation of abnormal, misfolded proteins [8], [9].

In the case of a larger group of amino acids however, the change in codon usage favoured translationally nonoptimal codons that lack matching isodecoder tRNA genes in mammals. The most likely explanation for this observation is that slowdown of translation at these codons favours co-translational folding [10], thereby reducing the chances of misfolding and aggregation of misfolded proteins in long-lived animals. As pointed out by Pechmann and Frydman, mRNA sequences are under selection to optimize the cotranslational folding of polypeptides [11]. According to the hypothesis of these authors nonoptimal codons slow translation to facilitate co-translational folding by allowing the nascent chain more time to develop native-like structure. This interpretation is in agreement with the well-established observation that aging is characterized by loss of proteostasis and the consequent accumulation of misfolded proteins.

In summary, our studies have shown that there are significant lifespan-dependent changes of codon usage that may contribute significantly to improved co-translational folding, resulting in a more balanced proteostasis and a lower rate of cellular aging in long-lived mammals.

2. Results and discussion

2.1. Impact of hypermutability on codon bias of mammalian species

It has been shown earlier that the normal distribution of the RSCU values of synonymous codons is dictated primarily by the GC content of the coding sequences of the genome [12], [13].

To monitor the impact of GC content of coding genomes on codon bias of mammalian species, we have downloaded the RSCU (Relative Synonymous Codon Usage) values of amino acids for placental mammals listed in the Codon Statistics Database (<http://codonstatsdb.unr.edu>, [14]. We limited our study to the 96 mammalian species for which lifespan data are available in the AnAge Database of Animal Aging and Longevity (<https://genomics.senescence.info/>, [15].

To determine the overall GC content of the coding sequences (CDSs) for each of the 96 mammalian species, we have downloaded the GC content data of all genes from the Gene Stats section of the Codon Statistics Database [14]. The GC contents of the coding genomes of these species were calculated by weighting the contribution of the genes according to their length (see **Methods**). To visualize the influence of the GC content of coding genomes on codon usage, we have plotted the RSCU values of amino acids as a function of the GC content of the coding genome of the various species, with major emphasis on amino acids Arg, Ser, Thr, Ala, Pro that have at least one codon that contains a CpG dinucleotide. As a control, we have calculated the same parameters for all amino acids, except Trp and Met that are encoded by single codons. We have also calculated the RSCU values of stop codons as a function of the GC content of the coding genome of the various species.

In harmony with earlier conclusions [12], [13], these analyses have confirmed that the RSCU values of synonymous codons are dictated primarily by the GC content of the coding sequences of the genome. In all cases, the RSCU values of synonymous codons with higher GC content increase, those with lower GC content decrease linearly with GC content of the coding sequences of mammalian species (See **Supplementary file 1** and **Supplementary file 2**).

As a typical example we show the results of the analysis of codons encoding glycine (Fig. 1A).

In the majority of these linear relationships, the absolute values of Pearson's r correlation coefficients and R-Square values (COD, Coefficient Of Determination) are very high, close to 1.0 (**Supplementary file 2**), indicating that there is a strong correlation between RSCU values of synonymous codons and the GC content of the coding genome and that the fitted lines explain most of the variability of the response data around its mean.

Synonymous codons with similar GC content may have markedly different RSCU values. A major reason for such differences is depletion of codons containing hypermutable sites. This point may be best illustrated by the case of codons containing CpG dinucleotides, such as the GCG codon of alanine. (Fig. 1B). As shown in this figure, although codons GCC and GCG have the same GC-content, the RSCU value of the codon GCG, containing a hypermutable CpG site, is strikingly lower than that of GCC at all GC-contents.

There are eight codons that contain a hypermutable CpG dinucleotide: CGN (Arg) and NCG (Ser, Ala, Thr and Pro). It has been shown earlier that hypermutability of CpG dinucleotides has a significant impact on codon bias of these amino acids in organisms with strongly methylated genomes [16], [4], [5].

Methylation of cytosines of CpG dinucleotides of the eight codons in the sense strand favors CG > TG mutations resulting in seven missense and one nonsense mutation [5]. Methylation of cytosines of CpG dinucleotides of these codons in the antisense strand leads to silent substitutions of the NCG codons of Thr, Pro, Ala and Ser, whereas substitutions of the four CGN codons of Arg lead to missense mutations. In other words, the four Arg codons of the CGN codon family are special among CpG containing codons in that C > T mutation of CpG of both the sense and antisense strand result in a change of the amino acid, whereas in the case of the NCG codons of Thr, Pro, Ala and Ser only the sense strand mutations alter the amino acid. It has been pointed out earlier that this difference of CpG bearing codons has significant evolutionary consequences for organisms with methylated genomes [16]. Thanks to purifying selection, NCA codons for Thr, Pro, Ala and Ser increase preferentially with stronger methylation, indicating that depression of NCG codons in strongly methylated genes occurs through silent C > T substitutions on the antisense strand. The CGN group of CpG bearing codons, which code for arginine evolve differently because C > T substitutions on both the sense and antisense strands produce amino acid changes. Dixon and coworkers have shown that arginine content is negatively correlated with gene body methylation, suggesting that the shift in arginine content is due to CpG hypermutability of these codons [16].

As shown in **Supplementary file 2**, the $RSCU_{50GC}$ values ($RSCU$ values calculated for 50% GC content) of CpG bearing codons of Ser, Ala, Pro and Thr are significantly lower than those of their synonymous codons with the same GC-content (Ala: GCC, 1.401 versus GCG, 0.267; Pro: CCC, 1.095 versus CCG, 0.269; Thr: ACC, 1.270 versus ACG, 0.268; Ser: TCC, 1.187 versus TCG, 0.206). In the case of Arg we see much less depletion of CpG bearing codons relative to their GC content equivalent (AGG, 1.354 versus, CGA, 0.734 and CGT, 0.554). These observations are in harmony with the notion that depletion of the CpG bearing NCG codons of Ser, Ala, Pro and Thr occurs by selectively neutral silent substitutions, whereas depletion of CGN codons due to hypermutability of CpG dinucleotides can occur only through nonsense or potentially detrimental missense substitutions.

It is noteworthy, however, that codon usage of CGC and CGG ($RSCU_{50GC}$ values: 0.895 and 0.992, respectively) is less affected by CpG-associated depletion than those of CGA and CGT ($RSCU_{50GC}$ values: 0.734 and 0.554, respectively). The difference in the depletion of these two sets of CGN codons lies in the fact that in the DNA double helix CGC and CGG form three C:G pairs, whereas CGA and CGT form only two C:G pairs, i.e. there is a marked difference in the stability of the double helices formed by these codons. It is highly relevant in this respect that CpG mutation rates produced by 5mC deamination have been shown to be highly dependent on local GC content as DNA melting is the rate-limiting step for 5mC deamination [17]. According to this interpretation the higher degree of depletion of CGA and CGT could be due to higher rates of 5mC deamination of these codons. Another factor that might contribute to the more moderate depletion of the GC-rich codons is that mismatch repair is known to be most efficient in regions of stable double helix [18].

Hypermutable of the stopogenic CGA codon has another important consequence. As discussed earlier, the fact that out of the codons that can give rise to a stop codon by single nucleotide substitution, CGA

is exceptional in as much as it can generate a TGA stop codon at a very high rate [4]. This fact explains why in animals with strongly methylated genomes the RSCU value of TGA is much higher than that of its compositional equivalent TAG [4]. This is also true for mammals included in the present analysis: the $RSCU_{50GC}$ values of TGA and TAG are 1.416 and 0.677, respectively (**Supplementary file 2**).

2.2. Evolution of longevity and changes in codon bias in mammalian species

To monitor changes in codon bias, for each species we have subtracted the RSCU values expected at the given GC content ($RSCU_{exp}$) from the actual RSCU values ($RSCU_{obs}$) and plotted these deviations as a function of the known lifespan (years) of the species (**Supplementary file 3**).

To minimize the influence of major taxon-specific biological differences on such a hypothetical correlation, we have first restricted our analyses to 19 rodent species then we have checked whether the correlations observed in the case of rodents are also detectable across a wider range of mammals. The justification for this approach is that some taxon-specific biological differences are known to have greater influence on mortality and lifespan than cellular aging. For example, bats have several unique features (powered flight, viral tolerance and disease resistance, hibernation. torpor) that contribute to longer lifespan explaining why most bats have lifespans more than fourfold of similar-sized placental mammals [19], [20]. In the next part of our work, we have examined whether the various lifespan-dependent changes in codon usage observed in the analysis of rodents are detectable when we analyze a broader spectrum of mammalian taxa with greater biological diversity.

2.2.1. Longevity and depletion of hypermutable CpG bearing codons in mammalian species

To test the assumption that negative selection of hypermutable CpG bearing codons could play a role in the evolution of longevity in mammals, we have monitored changes of codon usage of the codon families CGN (Arg) and NCG (Ser, Ala, Thr and Pro) as a function of the lifespan of mammalian species.,

In the case of CpG-bearing codons of the NCG group, there was a significant linear correlation between lifespan of rodents and depletion of CCG (Pearson's $r = -0.4436$), TCG (Pearson's $r = -0.5724$) and ACG (Pearson's $r = -0.6549$), suggesting that decrease in the number of these mutation hotspots favours longevity of rodents.

Whereas in rodents there was a significant correlation between lifespan and depletion of the CpG bearing codons CCG, TCG and ACG, no such depletion was evident from the analysis of a broader spectrum of mammals. This observation suggests that depletion of these mutation hotspots does not play a general role in the longevity of mammals.

In the case of the CGN codons, there was a significant linear correlation between lifespan of rodents and depletion of CGA (Pearson's $r = -0.6096$) and CGT (Pearson's $r = -0.6007$) and increased use of CGC

(Pearson's $r = 0.6438$) and CGG (Pearson's $r = 0.6151$). A possible explanation for these observations is that depletion of the CGA and CGT codons occurs primarily by silent CGA, CGT > CGC, CGG substitutions and that substitution of the CGA, CGT codons with synonymous CGC, CGG codons favours longevity of rodents.

The significant correlation between lifespan and depletion of the arginine codon, CGA observed in rodents was, however, also observed across the wider range of mammals (Pearson's $r = -0.4051$) concomitant with the increased use of CGG (Pearson's $r = 0.3681$). These observations suggest that depletion of the hypermutable stopogenic CGA codon occurs primarily by silent CGA > CGG substitution and that substitution of the stopogenic CGA codon with a synonymous nonstopogenic codon favours longevity of mammals.

2.2.2. Longevity and changes in codon bias of codons lacking hypermutable CpG sites

Synonymous codon usage is known to be under selective pressure to optimize translation efficiency, thus there is a general correlation between codon usage and the abundance of cognate tRNAs [21], [22]. Codon bias favouring the most abundant tRNAs is most extreme in highly expressed genes, providing a fitness advantage through increased translation efficiency and accuracy of protein synthesis [12].

In view of the fact that tRNA abundance and codon bias coevolve, in this section we discuss changes in codon preference in the light of differences in the abundance of cognate tRNAs in mammals.

As pointed out by B  niti  re and coworkers, tRNA gene copy number is a good predictor of tRNA abundance [7], therefore we have examined the possible impact of tRNA gene copy number on codon bias (**Supplementary file 2**). Information on the number of tRNA genes in Metazoa, in Mammalia and the group of mammalian species included in the present study have been compiled from data provided by B  niti  re and coworkers [7].

In Metazoa the 18 amino acids that are encoded by multiple synonymous codons form two major groups: (i) those whose synonymous codons are translated by more than one distinct isodecoder tRNAs; (ii) those for which all synonymous codons are translated by a single isodecoder tRNA [7]. The first group includes amino acids encoded by codon sextets (Leu, Arg and Ser), quartets (Val, Gly, Ala, Pro and Thr), a triplet (Ile) and NNG/NNA duets (Glu, Gln and Lys). NNG/NNA duets are unique in that each synonymous codon has a matching isodecoder tRNA, whereas in the case of amino acids with codon sextets, quartets and triplet at least one of the codons has no matching isodecoder tRNA.

The second group for which all synonymous codons are translated by a single isodecoder tRNA corresponds to the six amino acids encoded by NNC/NNT duets (Phe, Cys, Tyr, Asp, His, Asn).

2.2.2.1. Longevity and changes in codon usage in rodents

In the case of several NNC/NNT duets that have a single isodecoder tRNA, our analyses have revealed significant correlations between lifespan of rodents and changes of codon preference of Asn (Pearson's $r = \pm 0.7867$), Asp (Pearson's $r = \pm 0.6639$), Cys (Pearson's $r = \pm 0.5259$), Phe (Pearson's $r = \pm 0.7530$) and Tyr (Pearson's $r = \pm 0.7446$). As a typical example, we show the results of the analyses of the data for the two codons of phenylalanine (Fig. 3, **panels A and B**).

Interestingly, in the case of Asn, Asp, Cys, Phe and Tyr the change in codon usage favoured the use of the NNT codons that have no matching isodecoder tRNA gene in mammals and disfavoured the translationally optimal NNC codons that have matching isodecoder tRNA genes (**Supplementary file 2**). The most likely explanation for this observation is that slowdown of translation at some nonoptimal Asn, Asp, Cys, Phe, Tyr codons favors co-translational folding [10], [11], thereby reducing the chances of misfolding and aggregation of misfolded proteins in long-lived rodents.

Change of codon usage favouring translationally nonoptimal codons over optimal codons is not restricted to the set of NNC/NNT duets.

For example, in the case of Ala, it is the GCT codon that has the highest number of cognate tRNA genes, whereas the GCC codon lacks matching isodecoder tRNA genes. A significant correlation was detected between lifespan and reduced use of the optimal GCT codon (Pearson's $r = -0.6105$), whereas the use of the nonoptimal GCC codon showed a weak lifespan-dependent increase (Pearson's $r = 0.2447$). In the case of Arg, it is the CGC codon that lacks cognate tRNA genes, whereas the CGT codon is the one that has the highest number of cognate tRNA genes. A significant association was found between lifespan and reduced use of the optimal CGT codon (Pearson's $r = -0.6007$), whereas the use of the nonoptimal CGC codon showed a lifespan-dependent increase (Pearson's $r = 0.6438$). The CCC codon of Pro lacks isodecoder tRNA genes, whereas CCT has the highest number of tRNA genes. We identified a significant correlation between lifespan and reduced use of the optimal CCT codon (Pearson's $r = -0.3885$), whereas the use of the nonoptimal CCC codon showed a significant lifespan-dependent increase (Pearson's $r = 0.4932$).

AGT is one of the serine codons that lack isodecoder tRNA genes, whereas the TCT codon of serine has the highest number of tRNA genes. We have observed a significant correlation between lifespan of rodents and reduced use of the optimal TCT codon (Pearson's $r = -0.6309$), whereas the use of the nonoptimal AGT codon showed significant lifespan-dependent increase (Pearson's $r = 0.5532$).

Several other amino acids, however, showed an opposite phenomenon: the change of codon usage favoured the translationally optimal codons that have the highest number of matching isodecoder tRNA genes and disfavoured the use of codons that have no matching isodecoder tRNA gene in mammals (**Supplementary file 2**).

This phenomenon is most striking in the case of Ile (Fig. 4). Out of the three codons of this amino acid the ATT codon has the highest number of matching isodecoder tRNA genes, whereas the ATC codon has no matching isodecoder tRNA gene in mammals (**Supplementary file 2**). A significant negative

correlation was observed between lifespan of rodents (Fig. 4. **Panels A and B**).and reduced use of the nonoptimal ATC codon (Pearson's $r = -0.7885$), whereas the use of the optimal ATT codon showed a lifespan-dependent increase (Pearson's $r = 0.7590$).

As another example, we may mention the case of Leu, where the CTG codon has the highest number of matching isodecoder tRNA genes in mammals, whereas the CTC codon that has no matching isodecoder tRNA gene. We have found a significant correlation between lifespan of rodents and reduced use of the nonoptimal CTC codon (Pearson's $r = -0.7331$), whereas the use of the optimal CTG codon showed a lifespan-dependent increase (Pearson's $r = 0.5497$).

Evolution of longevity of rodents may also favour selection for translationally optimal Val codons. In the case of Val, the GTG codon has the highest number of matching isodecoder tRNA genes, whereas the GTC codon has no matching isodecoder tRNA gene in mammals. A significant inverse correlation was identified between lifespan and reduced use of the nonoptimal GTC codon (Pearson's $r = -0.7723$), whereas the use of the optimal GTG codon showed a lifespan-dependent increase (Pearson's $r = 0.5407$).

The NNG/NN A duets (Glu, Gln, Lys), where both synonymous codons have a comparable number of matching isodecoder tRNA genes do not fit into either of the two categories listed above: there is no clear distinction between optimal and nonoptimal codons. We observed no significant correlation between change of codon usage and lifespan of Gln. Long-lived rodents, however, slightly favoured the use of GAA codon of Glu over GAG codon (Pearson's $r = \pm 0.3372$). Since the number of isodecoder tRNA genes for GAG and GAA are similar, the slight shift of codon preference is unlikely to be explained by differences in their optimality for translation or cotranslational folding.

Surprisingly, in the case of Lys, there was a strong correlation between changes in codon preference and lifespan of rodents (Pearson's $r = \pm 0.736$). Since in most mammals the numbers of genes of isodecoder tRNA for AAG and AAA are similar the change of codon preference favouring the AAA codon over AAG, is unlikely to be explained by differences in their optimality for translation and folding of proteins.

In summary, our analyses have identified three distinct groups of amino acids that show lifespan-dependent changes of codon bias in rodents: 1) amino acids, where there is a correlation between increased use of nonoptimal codons and increased lifespan of rodents (e.g. Asn, Asp, Cys, Phe, Tyr, Ala, Arg, Pro, Ser); 2) amino acids, where there is a correlation between increased use of optimal codons and increased lifespan of rodents (e.g. Gly, Ile, Leu, Thr, Val); 3) amino acids, where changes in codon preference can not be explained by distinction of optimal and nonoptimal codons (Glu, Lys).

The most likely explanation for the increased use of translationally optimal codons in long-lived rodents is that this change ensures increased accuracy of translation [23] and the lower rate of mistranslation results in a more balanced proteostasis and a lower rate of cellular aging in long-lived rodents. This interpretation is in harmony with recent studies on rodent species with diverse lifespans that have shown that the fidelity of protein synthesis is a determining factor in the evolution of longevity [24], [6], [25], the basal rate of translation errors has shown strong negative correlation with maximum lifespan.

Increased use of translationally nonoptimal codons of Asn, Asp, Cys, Phe, Tyr, Ala, Arg, Pro, Ser in long-lived rodents was more surprising as nonoptimal codons may increase the chance of mistranslation, ribosome stalling and this can result in premature translation termination [6]. Use of nonoptimal codons, however, may have significant benefits if slowdown of translation at such codons, favours correct co-translational folding by allowing the nascent chain more time to develop native-like structure before continuing translation [11]. This hypothesis implicitly assumes that such a beneficial role of nonoptimal codons of protein-coding genes is „position-specific”: they work best if they occur in regions separating distinct folding units. It is noteworthy in this respect that Pechmann and Frydman have noted that hydrogen-bonded turns that frequently connect secondary structural elements (α -helices and β -strands) are strongly enriched in conserved nonoptimal codons [11]. This principle is also valid for linker regions that connect domains of multidomain proteins. As pointed out by Jacobson and Clark, positioning of nonoptimal codons in linker regions could facilitate a domain-by-domain, beads-on-a-string folding mechanism [26].

We suggest that correct co-translational folding of multidomain proteins in long-lived rodents is facilitated by increased use of nonoptimal codons in the linker regions connecting individual protein domains. Given that multidomain proteins account for a substantial fraction of animal proteomes [27] and that their co-translational folding is particularly challenging [28], the enrichment of nonoptimal codons in these regions may contribute substantially to rodent longevity.

2.2.2.2. Longevity and changes in codon usage across diverse mammalian taxa

In the subsequent phase of our work, we have analysed whether the lifespan- dependent changes in codon usage detected in rodents could likewise be identified in a broader spectrum of biologically diverse mammals.

Comparison of the correlations between lifespan and changes in codon usage in the analysis of mammals with those observed when the analysis was limited to rodents revealed that several correlations significant in rodents are obscured when we analyse a more diverse group of mammals. Several correlations remained significant but in the majority of cases Pearson's r values of the correlations were lower when the analysis included a wider group of mammalian taxa (**Supplementary file 2**).

The correlations between mammalian lifespan and changes in codon usage deviate most significantly from linearity in the case of mammalian species with extreme longevity. As shown in panels C and D of Figs. 3 and 4, the slopes of the linear changes in codon usage are the steepest in the range of relatively short-lived mammals (lifespan range of 1–40 years), less steep in the range of moderately long-lived mammals (lifespan range of 41–60 years) and practically zero in the case of extremely long-lived mammals (lifespan range above 60 years). The hyperbolic curves reflecting the correlation of changes in codon usage and lifespan suggest that although the change of codon usage contributes to the evolution of extended lifespan of mammals, extreme longevity of species such as humans and whales are not due

to even more extreme shifts in codon usage. The most likely explanation for this phenomenon is that during evolution of increased lifespan changes in codon usage saturate the sites critical for efficient translation or folding of multidomain proteins, reaching a point in long-lived mammals where lifespan becomes limited by processes other than protein misfolding and collapse of proteostasis.

Similarly to rodents, significant changes of codon usage were observed in the case of several NNC/NNT duets that have a single isodecoder tRNA. Our analyses have revealed a linear relationship between change of codon preference and lifespan in the case of Asn (Pearson's $r = \pm 0.4419$), Asp (Pearson's $r = \pm 0.5123$), Phe (Pearson's $r = \pm 0.5994$, see Fig. 3, panels C and D) and Tyr (Pearson's $r = \pm 0.4747$). Similarly to rodents, in all these cases the change in codon usage favoured the use of the nonoptimal NNT codon that has no matching isodecoder tRNA gene in mammals (**Supplementary file 2**). In the case of the NNC/NNT duets, Cys and His, however, we observed only very weak correlation between change of codon usage and lifespan. We have detected weaker but significant correlations between lifespan of mammals and reduced use of the optimal GCT codon of Ala and the optimal CCT codon of Pro. In contrast with the rodents-analysis, no significant correlation was found between mammalian lifespan and reduced use of optimal and increased use of nonoptimal codons of Ser and Arg.

Change of codon usage favouring the translationally optimal codons was far more restricted across mammals than in the case of rodents-analysis. We have observed no significant correlation between lifespan of mammals and reduced use of the nonoptimal and increased use of the optimal codons of Gly, Leu, Thr and Val. Only in the case of Ile did we observe a significant correlation between mammalian lifespan and reduced use of the nonoptimal ATC (Pearson's $r = -0.5485$ and increased use of the optimal the ATT codon (Pearson's $r = 0.5407$) (Fig. 4, **panels C and D**).

The results on the NNG/NNA duets (Glu, Gln, Lys) where each synonymous codon has a comparable number of matching isodecoder tRNA genes, were quite similar in the two analyses. In the case Gln we observed a weak correlation between change of codon usage and lifespan, long-lived mammals preferring the CAA over CAG. Long-lived mammals also favour the use of GAA codon of Glu over GAG codon (Pearson's $r = 0.3976$). Analyses of mammals also revealed a strong correlation between changes in codon preference of Lys, favoring the AAA codon over AAG, and lifespan (Pearson's $r = 0.5412$).

In summary, our analyses have shown that several of the significant lifespan-dependent changes in codon usage observed in rodents are also observed when we analyse a broader spectrum of mammals. Significantly, there is a strong correlation between increased use of nonoptimal codons and increased lifespan of mammals in the case of Asn, Asp, Phe, Tyr, Ala, and Pro whereas there is a strong correlation between increased use of optimal codon and mammalian lifespan only in the case of Ile. Since preference of translationally nonoptimal codons may be essential for correct co-translational folding [26] this observation suggests that correct protein folding is more critical for mammalian longevity than high rate of translation that would favour translationally optimal codons.

Since multidomain proteins constitute a very large fraction of the proteomes of animals and much of their organismic complexity is ensured by these proteins [27] their correct folding is of extreme

importance. The folding of multidomain proteins, however, differs from that of small single-domain proteins, which usually fold quickly and reversibly. Co-translational processes are important aspects of multi-domain protein folding: as the ribosome sequentially adds amino acids from N- to C-terminus, the growing nascent polypeptide begins to form tertiary structures, domain by domain [26], [28], [29]. We suggest that increased use of translationally nonoptimal codons in linker regions of multidomain proteins favours their correct co-translational folding and may thus play a critical role in the evolution of mammalian longevity.

Although our work focused on codon bias of synonymous sense codons, we note with interest that there is a significant correlation between mammalian lifespan and increased use of the TAA stop codon (Pearson's $r = 0.4333$) and decreased use of TAG (Pearson's $r = -0.4279$) and TGA (Pearson's $r = -0.2542$) as stop codons (**Supplementary file 2**). It is now well established that the three stop codons are not fully synonymous as they differ in efficiency to terminate translation [4]. TAA appears to be the most efficient, whereas the others permit significant readthrough due to misinterpretation of the stop codon. As we have discussed earlier [4], it has been debated for a long time whether readthrough of termination codons is detrimental (as it creates aberrant forms of proteins) or has adaptive value (as it provides a means for proteome expansion). Our observation that evolution of mammalian longevity favors the use of TAA that is the least prone to readthrough suggests that avoidance of the formation of aberrant forms of proteins is of primary importance for lower rates of cellular aging.

3. Conclusion

In the present work we have detected significant correlations between evolution of lifespan in mammals and changes in codon usage of the majority of synonymous codons.

For example, the hypermutable CpG bearing stopogenic arginine codon, CGA, is significantly more depleted in long-lived than short-lived mammals. The most likely explanation for this observation is that negative selection of this codon renders genes less vulnerable [5], [30], promoting the evolution of longevity.

We have also noted that there is a significant correlation between extended mammalian lifespan and increased use of the TAA stop codon (and decreased use of TAG and TGA as stop codons). Since TAA is the least prone to stop codon readthrough, this observation suggests that avoidance of the formation of abnormal proteins with aberrant C-terminal extensions is important for lower rates of cellular aging.

In addition to these special cases, our analyses have revealed remarkable correlations between evolution of longevity in mammals and changes in usage of the majority of synonymous codons.

In the case of some amino acids (e.g., Ile) the change in codon usage accompanying evolution of longevity favoured translationally optimal codons, i.e. those that match the most abundant isodecoder tRNAs. The most plausible explanation for this observation is that these translationally optimal codons

ensure higher accuracy (lower rate of mistranslation), thereby reducing the chances of the formation of abnormal proteins in long-lived animals [6], [7].

Unexpectedly, in the case of a much larger group of amino acids (e.g., Tyr, Phe, Asp, Asn), the change in codon usage accompanying evolution of longevity favoured translationally nonoptimal codons that lack matching isodecoder tRNA genes. In our view, the most likely explanation for this observation is that slowdown of translation at these codons favours co-translational folding as suggested earlier [10], [11], [26]. According to the hypothesis of these authors, nonoptimal codons slow translation and facilitate co-translational folding by allowing the nascent chain more time to develop native-like structure.

Co-translational processes are especially important for folding of multi-domain proteins as the growing polypeptide must form tertiary structures domain by domain [26], [28], [29]. We suggest that increased use of translationally nonoptimal codons in critical regions of proteins (such as linker regions of multidomain proteins) favours their correct co-translational folding and may thus play a critical role in the evolution of mammalian longevity (Fig. 5). This explanation is also consistent with the central role of proteostasis loss in aging, manifested by the accumulation of misfolded proteins.

4. Methods

4.1. Lifespan data of mammalian species

Lifespan information (maximum lifespan values) for mammalian species was retrieved from the AnAge Database of Animal Aging and Longevity (<https://genomics.senescence.info/>, [15]).

4.2. Codon usage and genomic GC content data of mammalian species

The relative synonymous codon usage (RSCU) values for the coding sequences of mammalian species were downloaded from the Codon Statistics Database (<http://codonstatsdb.unr.edu>, [14].) over the period December 3–13, 2024. We limited our study to the 96 mammalian species for which complete datasets were available in both the Codon Statistics Database and the AnAge Database.

To determine the overall GC content of the coding genome for the 96 mammalian species, we have downloaded the GC content data of genes from the Gene Stats section of the Codon Statistics Database (**Supplementary file 4**). The GC contents of the coding genomes of these species were calculated by weighting the contribution of the genes according to their length. For each gene of a given species the GC fraction values were multiplied by the number of codons present in the gene and the sum of these values of all genes were divided by the total number of codons of all genes.

To visualize the influence of the GC content of coding genomes on codon usage, we have plotted the RSCU values of amino acids as a function of the GC content of the Coding Genome of the various species.

4.3. Data analysis and statistical procedures

To monitor the possible correlations between changes in codon bias and lifespan we have calculated the RSCU values expected at the given GC content ($RSCU_{exp}$), based on the linear relationship between the RSCU and GC content. $RSCU_{exp}$ values were subtracted from the actual RSCU values ($RSCU_{obs}$), and we have plotted these deviations as a function of the known lifespan (years) of the species.

All datasets were processed and analyzed using OriginPro 2025. Regression analyses and Pearson's correlation computations were performed to assess the relationships between codon usage metrics, GC content and species lifespan. In the evaluation of the results of these analyses we have focused on strong correlations, i.e. cases where the absolute value of Pearson's correlation coefficient (r) was higher than 0.5. Cases where the absolute value of Pearson's correlation coefficient (r) was between 0.3 and 0.5 are discussed as moderate but significant correlations.

Declarations

Funding

Open access funding has been provided by HUN-REN Research Centre for Natural Sciences, Hungarian National Research, Development and Innovation Office, GINOP-2.3.2-15-2016-00001. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Conflict of interest

None declared.

Acknowledgments

KK, MT, LB and LP are supported by the GINOP-2.3.2-15-2016-00001 grant of the Hungarian National Research, Development and Innovation Office (NKFIH). We appreciate the tools of BioRender that were instrumental for the visualization of our findings.

Contributor Information

Krisztina Kerekes, Email: kerekes.krisztina@ttk.hu

Mária Trexler, Email: trexlermari@gmail.com

László Bányai, Email: banyai.laszlo@ttk.hu

László Patthy, Email: patthy.laszlo@ttk.hu

Krisztina Kerekes, Mária Trexler, László Bányai: Formal analysis, Validation, Investigation, Methodology, Writing—original draft, Writing—review and editing; László Patthy, Conceptualization, Supervision,

Funding acquisition, Validation, Methodology, Writing—original draft, Project administration, Writing—review and editing.

Data availability

All data generated or analysed during this study are included in the manuscript and supporting files.

References

1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nat.* **458** no. **7239**, 719–724. 10.1038/nature07943 (2009).
2. Szilard, L. On the nature of the aging process. *Proc. Natl. Acad. Sci. U S A* **45** no. (1), 30–45. 10.1073/pnas.45.1.30 (1959).
3. Morley, A. A. The somatic mutation theory of ageing. *Mutat. Res.* **338**, 19–23. 10.1016/0921-8734(95)00007-s (1995).
4. Trexler, M., Bányai, L., Kerekes, K. & Patthy, L. Evolution of termination codons of proteins and the TAG-TGA paradox. *Sci. Rep.* **13** no. (1), 14294. 10.1038/s41598-023-41410-z (2023).
5. Trexler, M., Bányai, L., Kerekes, K. & Patthy, L. Arginines of the CGN codon family are Achilles' heels of cancer genes. *Sci. Rep.* **14** no. (1), 11715. 10.1038/s41598-024-62553-7 (2024).
6. Ke, Z. et al. 2017 Translation fidelity coevolves with longevity. *Aging Cell.* **16** no 5: 988–993. 10.1111/accel.12628
7. Bénitière, F., Lefébure, T. & Duret, L. Variation in the fitness impact of translationally optimal codons among animals. *Genome Res.* **35** no. **3**, 446–458. 10.1101/gr.279837.124 (2025).
8. Taylor, R. C. & Dillin, A. Aging as an event of proteostasis collapse. *Cold Spring Harb Perspect Biol.* **3** 5: a004440. (2011). 10.1101/cshperspect.a004440
9. Cuanalo-Contreras, K., Mukherjee, A. & Soto, C. Role of protein misfolding and proteostasis deficiency in protein misfolding diseases and aging. *Int J Cell Biol.* 2013: 638083. (2013). 10.1155/2013/638083
10. Sherman, M. Y. & Qian, S. B. Less is more: improving proteostasis by translation slow down. *Trends Biochem. Sci.* **38** no. **12**, 585–591. 10.1016/j.tibs.2013.09.003 (2013).
11. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–243. 10.1038/nsmb.2466 (2013).
12. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12** no. (1), 32–42. 10.1038/nrg2899 (2011).
13. Trotta, E. Selective forces and mutational biases drive stop codon usage in the human genome: a comparison with sense codon usage. *BMC Genom.* **17**, 366. 10.1186/s12864-016-2692-4 (2016).
14. Subramanian, K., Payne, B., Feyertag, F. & Alvarez-Ponce, D. The Codon Statistics Database: A Database of Codon Usage Bias. *Mol. Biol. Evol.* **39** no, 8: msac157. 10.1093/molbev/msac157

(2022).

15. de Magalhães, J. P. et al. 2024 Human Ageing Genomic Resources: updates on key databases in ageing research. *Nucleic Acids Res.* **52**(D1):D900–D908. 10.1093/nar/gkad927
16. Dixon, G. B., Bay, L. K. & Matz, M. V. Evolutionary Consequences of DNA Methylation in a Basal Metazoan. *Mol. Biol. Evol.* **33** no. **9**, 2285–2293. 10.1093/molbev/msw100 (2016).
17. Fryxell, K. J. & Moon, W. J. CpG mutation rates in the human genome are highly dependent on local GC content. *Mol. Biol. Evol.* **22** no. **3**, 650–658. 10.1093/molbev/msi043 (2005).
18. Jones, M., Wagner, R. & Radman, M. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genet.* **115** no. **4**, 605–610. 10.1093/genetics/115.4.605 (1987).
19. Wilkinson, G. S. & Adams, D. M. Recurrent evolution of extreme longevity in bats. *Biol. Lett.* **15** no. **4**, 20180860. 10.1098/rsbl.2018.0860 (2019).
20. Morales, A. E. et al. 2025 Bat genomes illuminate adaptations to viral tolerance and disease resistance. *Nat.* **638** no **8050**: 449–458. 10.1038/s41586-024-08471-0
21. Bulmer, M. Coevolution of codon usage and transfer RNA abundance. *Nature*. Feb 19–25;325(6106):728 – 30. (1987). 10.1038/325728a0. PMID: 2434856.
22. Moriyama, E. N. & Powell, J. R. Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol.* ;45(5):514 – 23. (1997). 10.1007/pl00006256. PMID: 9342399.
23. Sun, M. & Zhang, J. Preferred synonymous codons are translated more accurately: Proteomic evidence, among-species variation, and mechanistic basis. *Sci. Adv.* **8** no. **27**, eabl9812. 10.1126/sciadv.abl9812 (2022).
24. Azpurua, J. et al. 2013 Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. *Proc Natl. Acad. Sci. U S A* **110** no. **43**: 17350–17355. 10.1073/pnas.1313473110
25. Ke, Z., Seluanov, A. & Gorbunova, V. Accurate translation is important for longevity. *Aging (Albany NY)*. **10** no. **3**: 297–298. (2018). 10.18632/aging.101398
26. Jacobson, G. N. & Clark, P. L. Quality over quantity: optimizing co-translational protein folding with non-'optimal' synonymous codons. *Curr. Opin. Struct. Biol.* **38**, 102–110. 10.1016/j.sbi.2016.06.002 (2016).
27. Tordai, H., Nagy, A., Farkas, K., Bányai, L. & Patthy, L. Modules, multidomain proteins and organismic complexity. *FEBS J.* **272** no. **19**, 5064–5078. 10.1111/j.1742-4658.2005.04917.x (2005).
28. Rajasekaran, N. & Kaiser, C. M. Co-Translational Folding of Multi-Domain Proteins. *Front. Mol. Biosci.* **9**, 869027. 10.3389/fmolb.2022.869027 (2022).
29. Rajasekaran, N. & Kaiser, C. M. Navigating the complexities of multi-domain protein folding. *Curr. Opin. Struct. Biol.* **86**, 102790. 10.1016/j.sbi.2024.102790 (2024).
30. Di Fraia, D. et al. 2025 Altered translation elongation contributes to key hallmarks of aging in the killifish brain. *Sci.* **389** no **6759**: eadk3079. 10.1126/science.adk3079

Figures

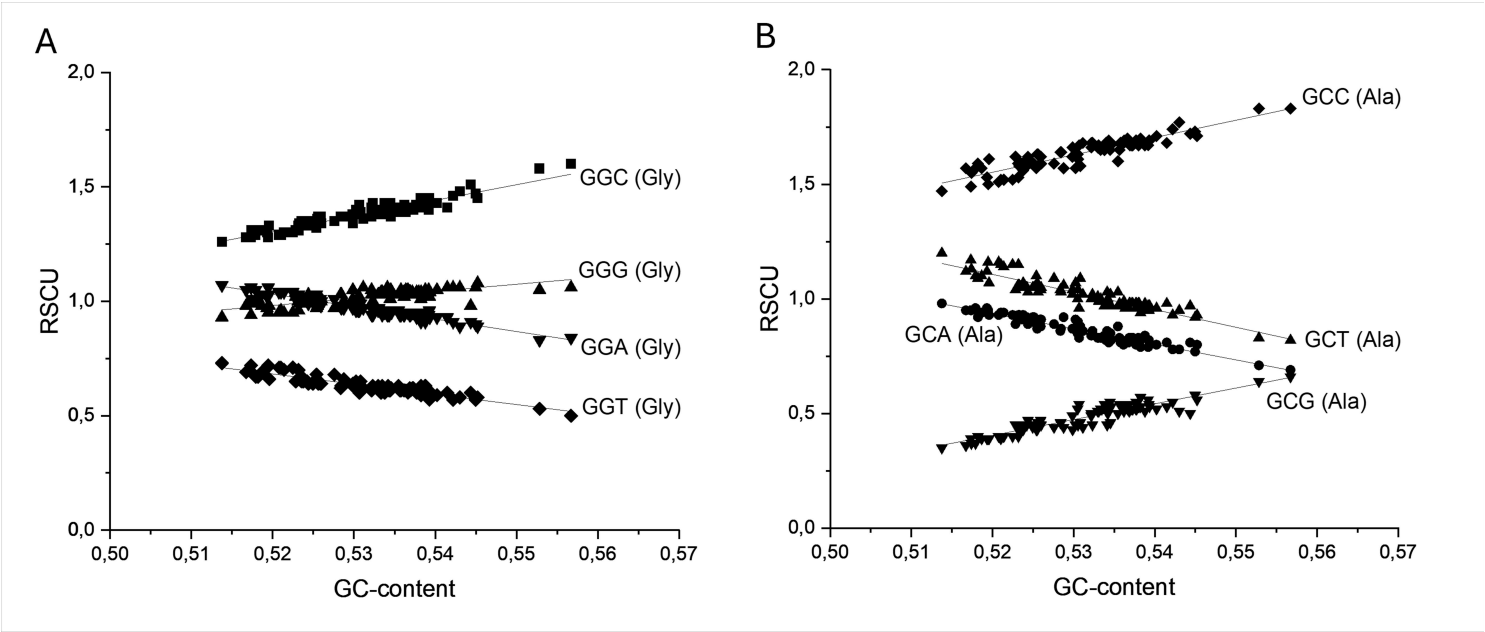


Figure 1

Relative synonymous codon usage (RSCU) of the four glycine (A) and the four alanine (B) codons plotted against the GC content of the coding genomes of the 96 mammalian species included in the present analysis.

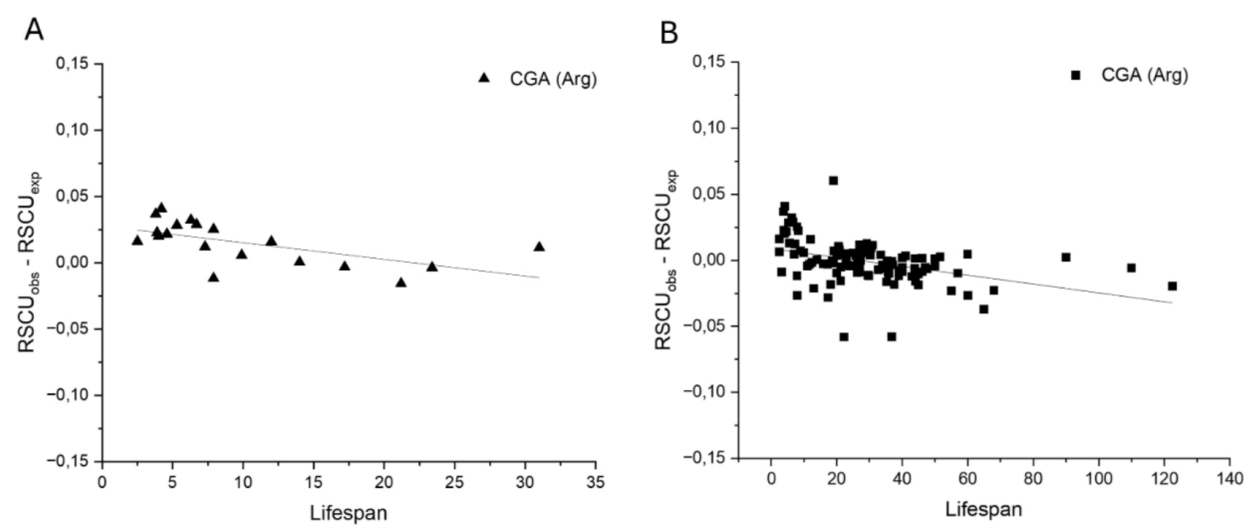


Figure 2

Correlation between lifespan and CGA codon usage bias in the species included in the present study, comprising 19 rodent species (panel A) and 96 mammalian species (panel B).

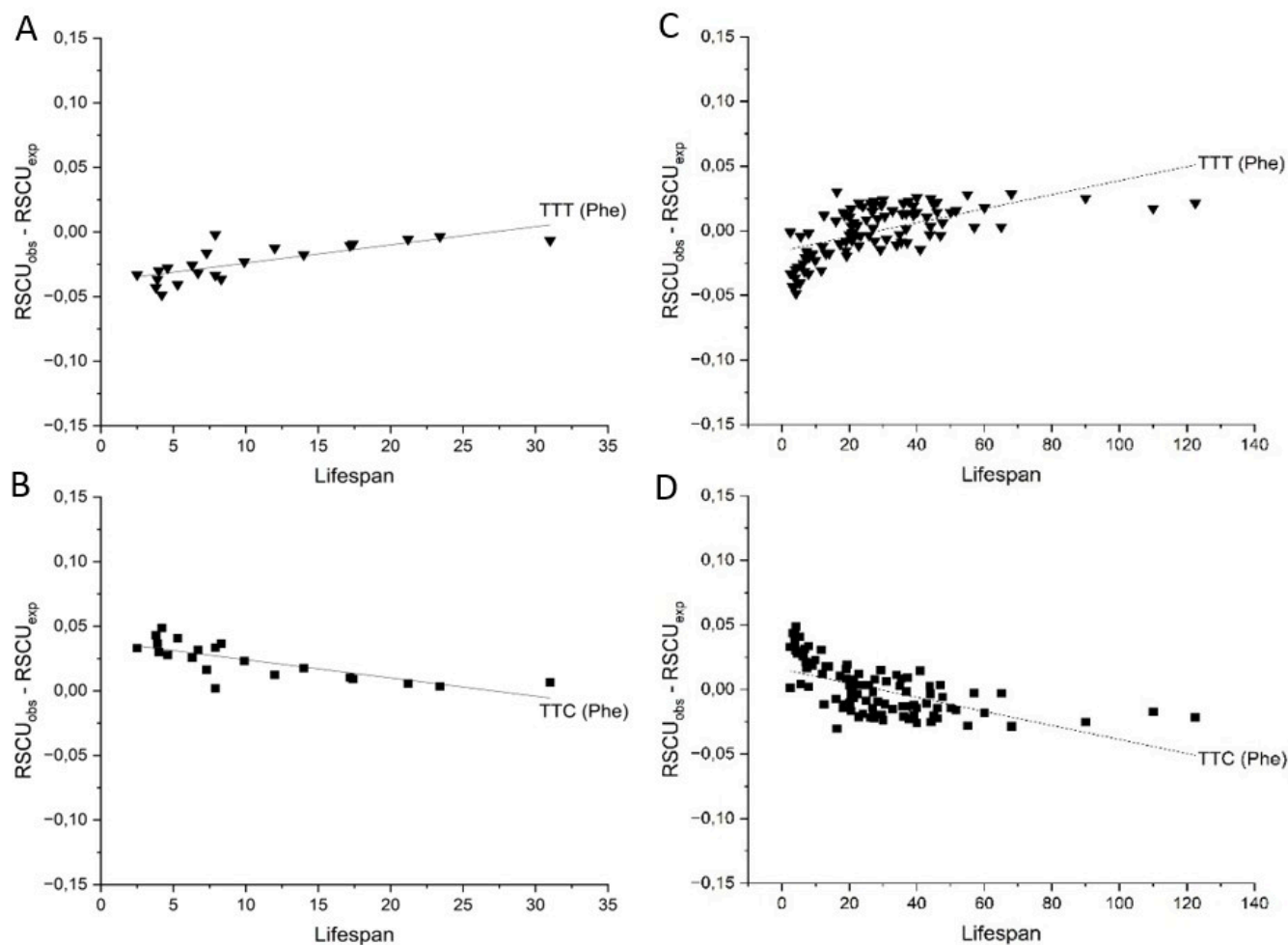


Figure 3

Correlation between lifespan of mammalian species and changes in codon preference of phenylalanine (Phe). Panels A and B show data for the 19 rodent species, Panels C and D show data for the 96 mammalian species included in the present analysis.

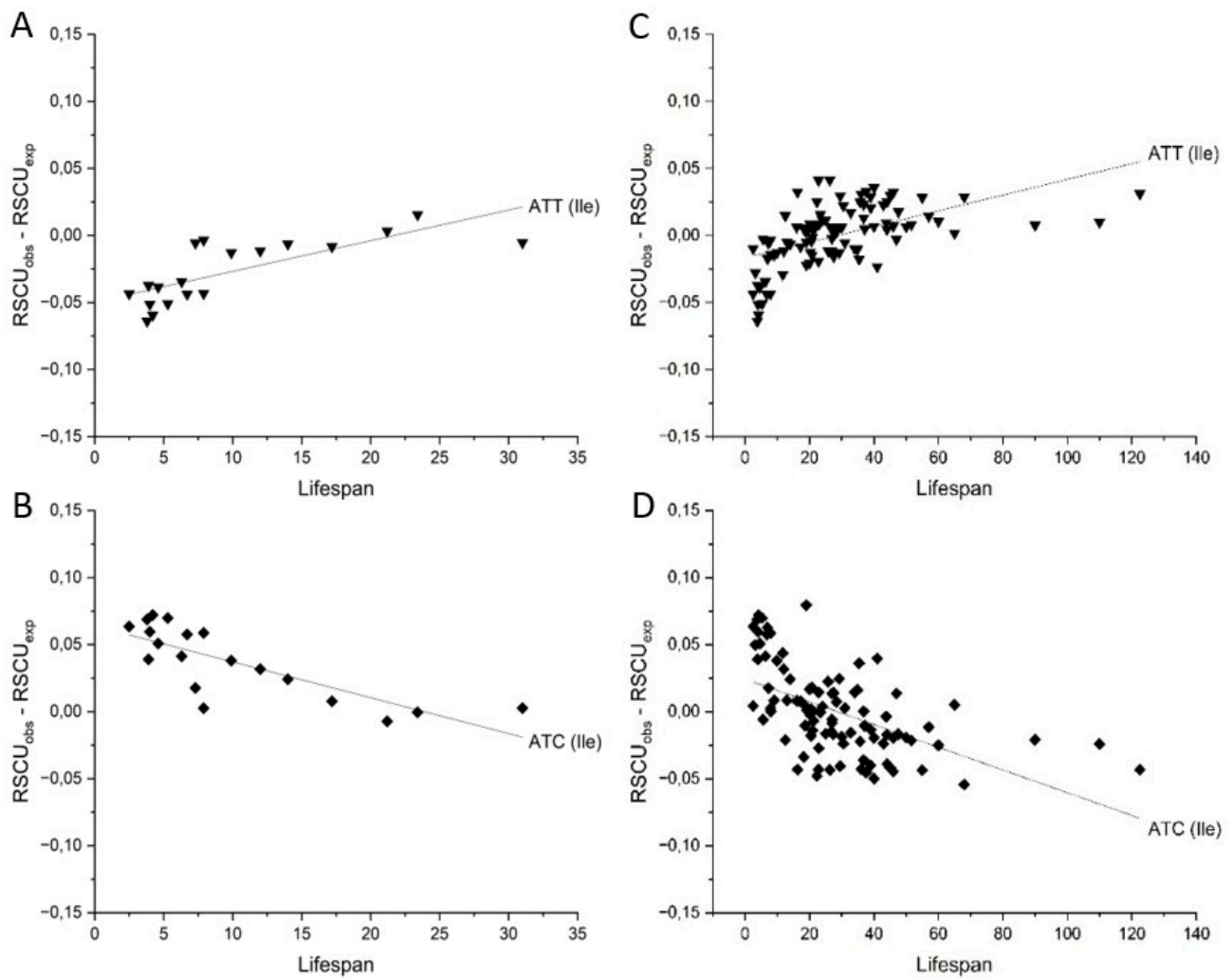


Figure 4

Correlation between changes in codon preference of isoleucine (Ile) and lifespan of mammalian species. Panels A and B show data for the 19 rodent species, Panels C and D show data for the 96 mammalian species included in the present analysis.

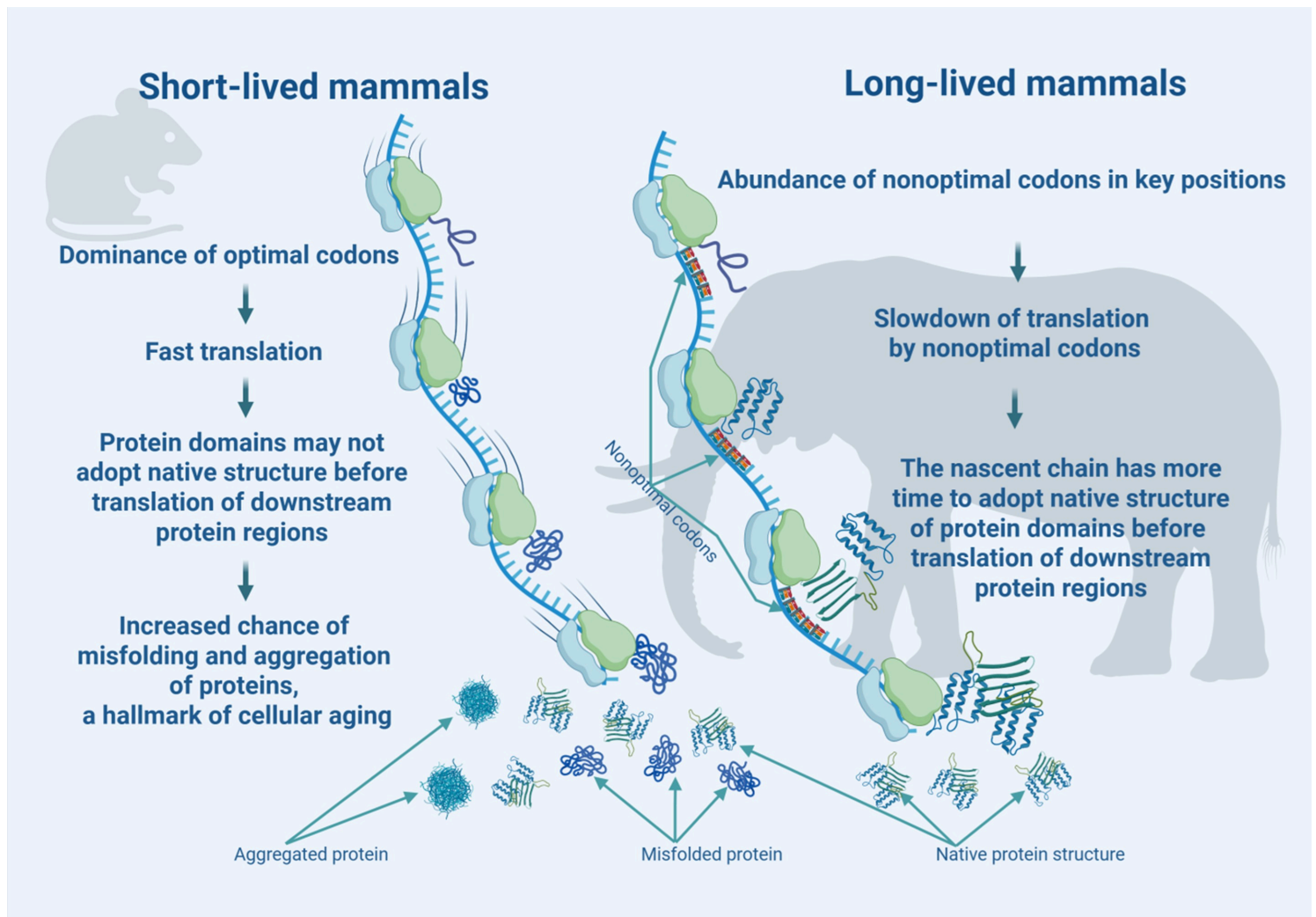


Figure 5

Proposed model for the effects of optimal and nonoptimal codons on folding of proteins and longevity. Created with BioRender.com. In short-lived mammals translation is fast due to the dominance of optimal codons that match the most abundant isodecoder tRNAs, therefore the protein domains may not adopt their stable native structure before translation of downstream protein regions. High speed of translation may increase the chance of misfolding and aggregation of misfolded proteins, a hallmark of cellular aging. In long-lived mammals nonoptimal codons replace optimal codons in critical regions (e.g., linker regions of multidomain proteins). Slowdown of translation at these codons favours co-translational folding by allowing the nascent chain more time to develop native-like structure before translation of downstream protein regions. Thus, replacement of optimal codons by nonoptimal codons in critical regions promotes co-translational folding and decreases the chance of misfolding and aggregation of misfolded proteins.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.pdf](#)