RECODING OF TRANSLATION IN TURTLE MITOCHONDRIAL GENOMES: PROGRAMMED FRAMESHIFT MUTATIONS AND EVIDENCE OF A MODIFIED GENETIC CODE

by

Robert David Russell
B.Sc.H. (Biology), Queen's University, 2003
B.A. (Computer Science), Queen's University, 2003

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department
of
Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Summer 2006

APPROVAL

NAME: Robert David Russell

DEGREE: Master of Science (Molecular Biology and Biochemistry)

TITLE OF THESIS: Recoding of Translation in Turtle Mitochondrial Genomes: Programmed Frameshift Mutations and Evidence of a Modified Genetic Code

EXAMINING COMMITTEE:

CHAIR: Dr. Melanie O'Neill
Assistant Professor, Dept. of Chemistry

_____
Dr. Andrew T. Beckenbach
Senior Supervisor
Professor, Dept. of Biological Sciences

_____
Dr. David L. Baillie
Supervisory Committee Member
Professor, Dept. of Molecular Biology and Biochemistry

_____
Dr. Peter J. Unrau
Supervisory Committee Member
Assistant Professor, Dept. of Molecular Biology and Biochemistry

_____
Dr. Barry M. Honda
Public Examiner
Professor, Dept. of Molecular Biology and Biochemistry

Date Approved: August 2, 2006

ABSTRACT

A +1 frameshift insertion has been documented in the mitochondrial gene *nad3* in some birds and reptiles. By sequencing the polyadenylated mRNA in the chicken (*Gallus gallus*), we have shown that the extra nucleotide is transcribed and is present in the mature mRNA. Evidence from other animal mitochondrial genomes has led us to hypothesize that certain mitochondrial translation systems have the ability to tolerate frameshift insertions using programmed translational frameshifting. To investigate this, we sequenced the mitochondrial genome of the red-eared slider turtle (*Trachemys scripta*), where both the common *nad3* frameshift insertion and a novel site in *nad4l* were found. Sequencing the region surrounding the insertion in *nad3* in a number of other turtles and tortoises revealed general mitochondrial +1 programmed frameshift site features as well as the apparent redefinition of a stop codon in Parker's sideneck turtle (*Macrochelodina parkeri*), the first known example of this in vertebrate mitochondria.

Keywords: programmed translational frameshifting, mitochondrial genetic code, *Testudines*, recoding, codon redefinition

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped me to get to this point – if you are not listed below, I apologize, and I can only hope that you know I appreciate all you have done for me.

First, I would like to thank my senior supervisor, Dr. Andy Beckenbach, for providing me with this opportunity, for helping develop what turned out to be a very interesting project, and for supporting me without concession throughout my research. My supervisory committee of Dr. Dave Baillie and Dr. Peter Unrau also deserve much credit for their guidance and valuable insights into the project.

A very special thanks is due to Dr. Jim Stewart and Andrea Scouras, who not only welcomed me warmly into the lab, but also displayed an inexhaustible supply of patience in teaching me everything I needed to know. I would also like to thank my other labmates over the course of my studies, Mike Vankoeveringe and Jeremy Saunier, for their help and advice along the way.

I owe much gratitude toward Eric Holt at the Empire of the Turtle and Paul Springate at the Rainforest Reptile Refuge for providing invaluable samples, without which this project would have never got off the ground.

To my family and friends I would like to express my sincere thanks and appreciation for their unwavering support, especially to my parents, who are, in a word, amazing.

Finally, to Poet, thank you, for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

We have discovered a number of unique coding events in the mitochondria of various turtles and tortoises (chelonians). Here we report the complete mitochondrial genome sequence of the red-eared slider, *Trachemys scripta*. This genome includes both a conserved programmed translational frameshifting site, as well as a novel one not previously documented. To characterize the conserved frameshift site in turtles, we have sequenced the surrounding region of the mitochondrial genome in a number of different species. During this we discovered another interesting coding event, an apparent case of codon redefinition in *Macrochelodina* (formerly *Chelodina*) *parkeri* (Parker's side-necked turtle). Through a variety of techniques, we have analysed each of these regions in an effort to understand further the translational mechanisms that are involved. The following introduction will provide a background in recoding phenomena with an emphasis on +1 programmed translational frameshifting.

Recoding

Normal translation

The cellular mechanisms for standard translation that convert messenger RNA (mRNA) to polypeptides are well understood. The ribosome, a large ribonucleoprotein complex that catalyzes the translation of protein from mRNA, reads sequentially displayed nucleotide triplets, or codons. The amino-acyl tRNA (aa-tRNA) with a

complementary anticodon is brought to the aa-tRNA binding site (A-site) by an elongation factor (EF-Tu). Recognition of the codon by the tRNA and selection of the proper aa-tRNA occurs at this site. Conventionally, the ribosomal reading frame, or 0-frame, is established by a start or initiation codon, and codons are read in the open reading frame (ORF) until a stop codon is encountered. This process is relatively accurate, making an estimated 5×10^{-4} mistakes per amino acid in *Escherichia coli* (Parker, 1989). Beyond the mechanisms of normal translation however, there are also other unique, non-standard mechanisms employed by various organisms in all kingdoms of life that produce unexpected translation results at much higher frequencies than would be expected from baseline translational errors. These mechanisms cause deviations from the standard code in order to produce specific protein products. In essence, the mRNA is "re-coded" to produce a protein that differs from that predicted by the standard rules of translation.

Non-standard translation events

One of the first discoveries leading to the idea that there is something beyond the standard rules of translation operating in some organisms was a phage protein Qβ that required UGA, normally a stop codon, to be translated as tryptophan (Weiner and Weber, 1973). Other examples followed shortly after, until it was obvious that there are frequent deviations from the standard code in all kingdoms of life. Gesteland and Atkins first suggested the term 'recoding' in 1996 to encompass all the events during translation and transcription that do not conform to the standard rules of decoding. This includes the redefinition of codons – for example, the decoding UGA and UAG stop codons as the 21st amino acid selenocysteine, modified tRNAs charged with non-standard amino acids,

as well as programmed frameshifts, where the ribosome changes reading frames at certain sites. Frameshifting is a unique class of recoding event, as it produces a protein product that is completely different from the 0-frame predicted product downstream from the shift site.

Gesteland and Atkins define these dynamic reprogramming events with four specific criteria: 1 – the change from the normal coding rules occurs only at specific sites; 2 – the event occurs in competition with regular translation; 3 – the event occurs on mRNA; and 4 – there are signals within that mRNA that stimulate these events. As will be shown here, all of these conditions are met by the recoding events found in birds and turtles, including the two novel sites revealed in this study.

Programmed translational frameshifting

Insertion and deletion mutations and pseudogenes

Single nucleotide insertions and deletions in protein coding genes generally produce aberrant proteins, as the disruption in the reading frame causes the entire downstream sequence to be read out of frame. These proteins are also often truncated prematurely by downstream nonsense codons that are revealed in the new reading frame. In combination, these two problems typically either reduce or completely eliminate the ability of the protein to function as intended. In fact, the presence of these stop codons in the off-frames has inspired the hidden stop theory of mutational suppression, which argues that stop codons in reading frames other than the 0-frame are selected for to prevent indels (Seligmann and Pollock, 2004). Regardless, protein coding genes that

require more than one reading frame are usually classified as pseudogenes (for review, see D'Errico *et al.*, 2004), though, as we will see, this is not always the case.

The term pseudogene has varying definitions, but most commonly refers to genes that are assumed to be non-functional as a result of either some missing element(s) or from reading frame disruption. The problem with this definition is that occasionally, genes with coding sequence over more than one reading frame remain fully functional. The proviso is that they require a non-standard decoding event to take place in order to produce a functional protein product. In the case of a gene spanning more than one frame, a shift is required where the two frames join. If the sequence itself facilitates this shift, it is called a programmed translational frameshift.

Use of programmed translational frameshifting

Occasionally though, a single nucleotide indel occurs where it can be tolerated, and the gene can be translated completely in what was the original reading frame. This requires that the ribosome is somehow able to shift frames at this specific site at a high frequency. Though this phenomenon occurs rarely, it is found in all kingdoms of life, from bacteria to higher vertebrates. Depending on the sequence context and other factors, the shift in reading frame at a programmed frameshift site occurs at varying frequencies and is often in response to certain cellular conditions. In fact, some known examples of frameshifts are crucial to maintaining levels of their gene product through feedback mechanisms. Others allow organisms to produce multiple proteins from the same gene, while others hold no known importance and may persist only as a result of fortuitous mutational positioning.

General programmed translational frameshift characteristics

In bacteria, the chance that the ribosome will shift frames on any given codon is about 1 X 10^{-4} or less (Kurland, 1992b). When we compare this rate to certain sequences shown to frameshift up to 80% of the time at certain sites (Farabaugh, 2000), we realize the potential impact of such sites in protein production. The term programmed translational frameshift refers to the elements within certain mRNA sequences that are prone to frameshifting and mechanisms that act on these elements to increase the efficiency of the shift.

The two major classes of programmed frameshifting, +1 and -1 shifts, each have their own set of frameshifting stimulators. In most known examples of frameshifting sequences, certain elements have been shown to have profound effects on the level of frameshifting. Seemingly minor changes to these sites can completely eliminate all observable frameshifting. These elements include codons that are prone to ribosomal slippage, stimulatory RNA structures such as stem loops and pseudoknots, and the use of rare codons and stop codons which are thought to induce a stall in translation facilitating a shift in reading frame by the ribosome. There are also certain contextual sequences in some programmed frameshift examples with as yet unknown function. While there appear to be nearly as many ways that programmed translational frameshifting can be induced as there are examples of these phenomena, there are certain characteristics that are common across a number of different examples.

Brief overview of relevant programmed translational frameshift examples

The *gag* and *pol* polyproteins in Rous Sarcoma Virus that require a -1 frameshift to produce both proteins from the same gene were the first of many viral examples of programmed translational frameshifting (Jacks and Varmus, 1985). Ty1 and later Ty3 elements in yeast were another early frameshift example, again shifting between the *gag* and *pol* genes, the difference being a shift not to the -1-frame but to the +1 (Mellor *et al.*, 1985, Clare and Farabaugh, 1985, Wilson *et al.*, 1986). To cement these theories of recoding, an *E. coli* gene, release factor 2 (*prfB*), was discovered to contain an incomplete in-frame reading frame, and it was only through a +1 frameshifting event that the functional protein could be produced (Craigen *et al.*, 1985). These original discoveries have become the classical models for the mechanisms of programmed translational frameshifting to which all new examples are compared.

-1 Frameshifting

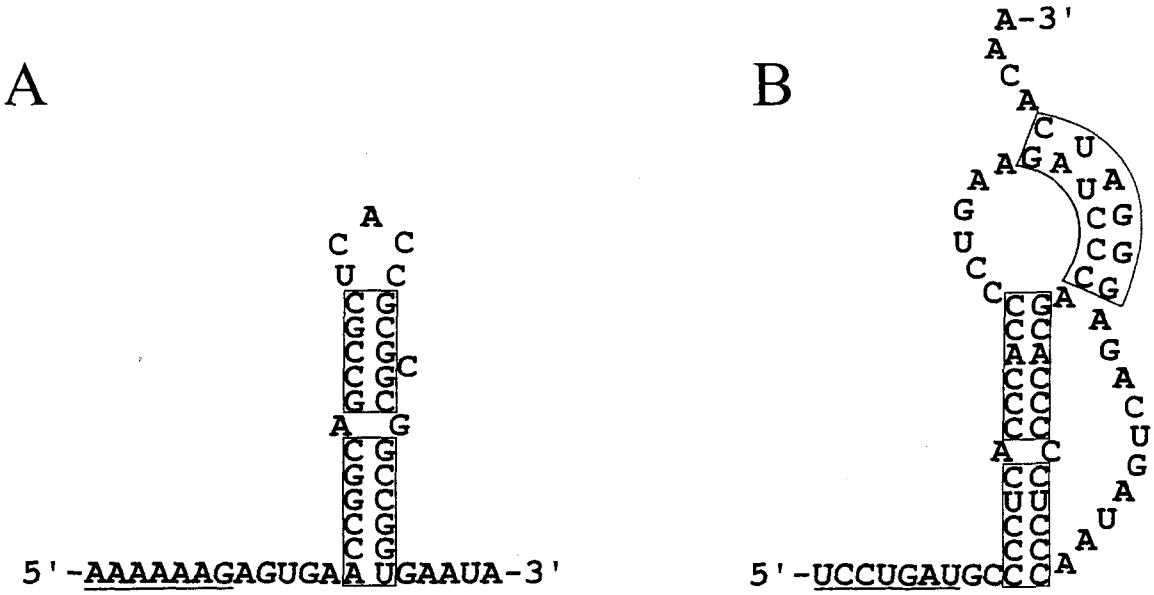
Frameshift events that move in the -1 direction relative to the 0-frame are generally composed of three distinct elements. From 5' – 3', these are nucleotide sequence that permits the ribosome to slip to the -1 frame, a spacer region, and an RNA structural element, which is generally a pseudoknot (Plant and Dinman, 2005). Though there are other sequences that can be used, in viral frameshifts the so-called "slippery sequence" usually takes the form of N NNW WWH, where NNN is any run of three of the same nucleotide, WWW is three residues of either A or U, and H is any of A, C or U. A pause by the ribosome, likely stimulated by the structural element (Tu *et al.*, 1992, Somogyi *et al.*, 1993), allows the sequence to shift from reading N NNW WWH to read NNN WWW H, and translation continues in the -1 frame. The translational pause happens while the A-

and P-site of the ribosome are in contact with the slippery heptamer sequence in the zero-frame.

One of the first examples of -1 programmed frameshifting observed was the *dnaX* gene in *E. coli*, encoding a DNA polymerase (Tsuchihashi and Kornberg, 1990, Flower and McHenry, 1990, Blinkowa and Walker, 1990). The sequence here is A AAA AAG which is shifted back a nucleotide to be read as AAA AAA G. A stem-loop structure 3' of the frameshift heptamer is thought to stall translation and initiate frameshifting (Tsuchihashi, 1991) (Figure 1-A). Though energetically similar pseudoknots and stem loop structures cause equivalent levels of ribosomal pausing, when compared, pseudoknots have been found to promote higher levels of -1 frameshifting (Kontos *et al.*, 2001). The final feature of the *dnaX* frameshift is a 5' Shine-Dalgarno sequence 10 bases from the frameshift site also shown to greatly stimulate frameshifting (Larsen *et al.*, 1994). Complementarity between this site and ribosomal RNA may cause increased frameshifting frequency. Although -1 frameshifting has not been found in any animal mitochondria genomes, it is mentioned here to introduce certain general frameshift site features such as mRNA secondary structure and stimulatory sequences, both of which are found in many +1 frameshifting events.

Figure 1 **Two examples of frameshift-stimulating RNA secondary structure.**

A. The stem-loop structure used in the *dnaX* -1 programmed translational frameshift. Stems of the hairpin are highlighted with rectangular outlines and the frameshift heptamer A AAA AAG is underlined. Partially stimulated by the stem-loop, this sequence shifts -1 to be read as AAA AAA G. B. The pseudoknot that is present 3 nucleotides downstream of the antizyme +1 frameshift site is represented here. Again, the frameshift heptamer is underlined, here the pseudoknot aids in the stimulation of a shift of the ribosome from reading UCC UGA U to read UCC U GAU (underlined). Complementary regions of the pseudoknot are outlined. For reasons unknown, 3' pseudoknots can stimulate both +1, as is the case here, and -1 frameshifting, as in many viral examples.



E. coli dnaX stem-loop

Human antizyme 1 pseudoknot

+1 Frameshifting

Frameshifting over an extra nucleotide, or +1 frameshifting, is the only type of frameshift recoding found in animal mitochondria to date, and is the only type seen in mammals. Like -1 frameshifting, a +1 frameshift is often stimulated by a pause in translation. In this case however, the pause is generally the result of a stop codon or rarely used codon as the last zero frame codon in the A-site, with a common codon present in the +1 frame. Rare codons have been found to profoundly influence frameshifting in *E. coli* (Weiss and Gallant, 1983), though a rare codon alone is often not

enough to stimulate the production of enough functional product of a required protein. Another characteristic often found in +1 frameshifting is an ability of the P-site tRNA anticodon to re-pair well in the +1 frame. Examples of +1 frameshifting occur in a range of organisms, including bacteria, yeast, and vertebrates (for a more in depth review see Baranov *et al.*, 2001 or Farabaugh, 1996a), and may or may not have regulatory function. Here I will go into some detail about three classical +1 frameshift sites and their proposed mechanisms (summarized in Table 1), as each has similarities to the mechanisms employed in the various animal mitochondrial frameshifts.

Table 1 **Summary of programmed translational frameshifting examples.**

The five frameshift examples given in the introduction are summarized here. The sequence over which translation is thought to shift frames is provided, spaced as both the original 0-frame translation and the frameshifted translation. Frameshift stimulators, such as RNA structure or rare codons, are also listed. Finally, the proposed mechanism by which the ribosome shifts frames is noted.

Gene	Organism	Sequence in the 0-frame	Shifted Sequence	Stimulators	Frameshift Mechanism
DNA Polymerase (*dnaX*)	*E. coli*	A AAA AAG	AAA AAA G	Slippery heptamer, 3' stem-loop	-1 Re-pairing of P-site and A-site tRNAs
Release factor 2 (*prfB*)	*E. coli*, other bacteria	CUU UGA C	CUU U GAC	Slippery codon, stop codon, 5' Shine-Dalgarno sequence, *prfB* concentration	+1 Re-pairing by the P-site leucine tRNA
Ty1 Element (*gag-pol*)	*S. cerevisiae*	CUU AGG C	CUU A GGC	Slippery codon, rare codon in A-site	+1 Re-pairing by the P-site tRNA
Ty3 Element (*gag-pol*)	*S. cerevisiae*	GCG AGU U	GCG A GUU	Slippery codon, rare codon in A-site	Occlusion of 1st position of A-site
Antizyme (*oaz*)	Higher eukaryotes	UCC UGA U (in vertebrates)	UCC U GAU	Slippery codon, stop codon, polyamine levels, 3' pseudoknot, 5' 50 nucleotide sequence	Occlusion of 1st position of A-site

Three major +1 programmed translational frameshift examples

Release factor 2

The gene for release factor 2 in *E. coli*, *prfB*, requires a +1 frameshift early in translation to produce a complete *prfB* protein (Figure 2). Through amino acid sequence and mRNA comparisons, the ribosome has been shown to shift to the +1 frame over the sequence CUU UGA C (Craigen *et al.*, 1985). Here the UGA stop codon in the 0-frame is thought to initiate the stall. After a successful frameshift, translation continues beyond it in the +1 frame, in effect reading the sequence as CUU U GAC. Whether or not the frameshift occurs is dependent on *prfB* concentration, so the frameshift is a feedback mechanism regulating *prfB* levels.

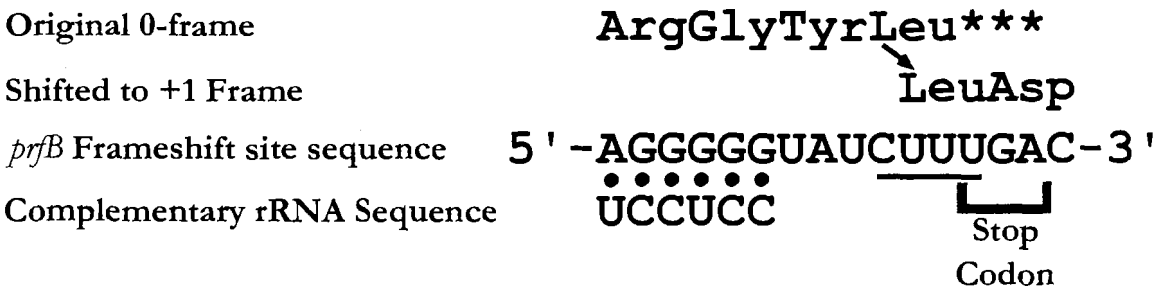
The CUU codon in *prfB* is an example of a slippery codon, and is used often as the last in-frame triplet decoded in other +1 frameshift sites. Slippery codons are thought to be key elements in frameshifting, allowing the mRNA to slide within the ribosome complex after the hydrogen bonds between the tRNA anticodon:codon pairs are severed when the codon is in the P-site (Weiss and Gallant, 1983, Tsuchihashi, 1991). Bonding is then re-establishing in the +1 frame. In *prfB*, tRNA-Leu (CUN) (anticodon GAG) base pairs with CUU, a shift prone codon (Curran, 1993), in the 0-frame (Figure 3). The ribosomal complex (specifically tRNA-Leu (CUN)) is able to shift frames from the CUU in the 0-frame to UUU in the +1 so long as there is a sufficient delay in the recognition of the UGA stop codon by *prfB*. The C after a UGA stop has been shown to reduce termination efficiency, possibly by delaying recognition of UGA by release factor (Poole *et al.*, 1995). Changing the leucine codon CUU or the C after the UGA stop codon at

position 26 decreases frameshifting frequency. Use of a leucine codon as the last in-frame codon in this way is present in many animal mitochondrial frameshift sites.

The final element of the *prfB* frameshift site is a Shine-Dalgarno sequence that also stimulates +1 shifting (Figure 2). This short sequence, 3 nucleotides 5' of the shift site, has been shown to base pair with the ribosome in a fashion similar to Shine-Dalgarno pairing with nucleotides 5' of the AUG start codon at the initiation of translation (Weiss *et al.*, 1987). This interaction may create a pause in translation (Gesteland and Atkins, 1996), or it may have some other more primary interaction with the ribosome or mRNA template strand, such as changing the conformation of one or the other.

Figure 2 ***E. coli prfB*** **frameshift site.**

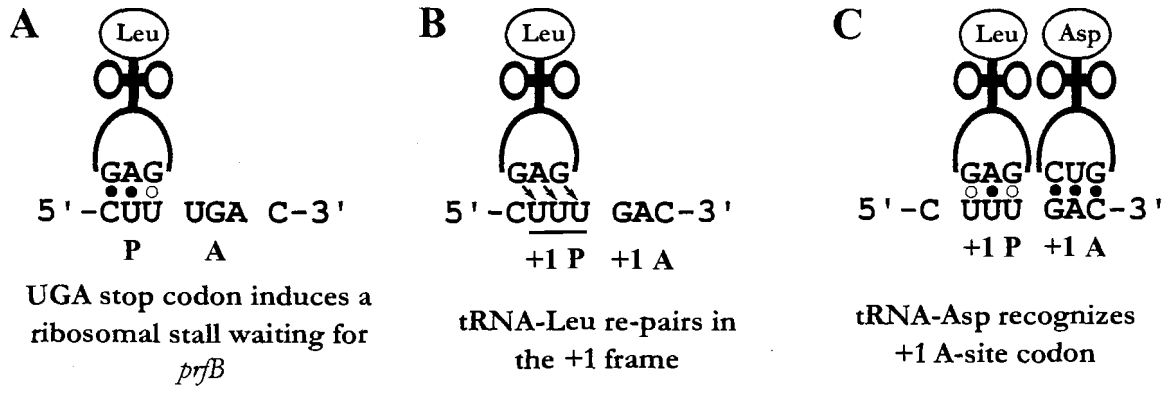
The frameshift in *prfB* occurs over a leucine codon, often present in +1 programmed translational frameshifts. Translation in the original frame reveals a UGA stop codon immediately following the frameshift site. The shift occurs over the CUU UGA C heptamer, where tRNA-Leu (CUN) in the P-site is believed to release from the CUU codon and re-pair in the +1 frame with UUU. Translation then continues downstream in the +1 frame. The upstream Shine-Dalgarno sequence and the complementary ribosomal RNA are shown. The positioning of the Shine-Dalgarno sequencing is crucial to frameshifting. Moving this site just one nucleotide upstream reduces frameshifting 17-fold (Weiss *et al.*, 1987).



Adapted from Farabaugh, 1996b.

Figure 3 **Proposed *prfB* frameshift mechanism.**

(A) In the initial step of the frameshift, the charged leucine tRNA recognizes the CUU codon in the 0-frame. Delay in recognition of the UGA stop codon by functional *prfB* protein causes the ribosome to stall. The frameshift heptamer is illustrated with schematic tRNAs bound with their cognate amino acid. Codon:anticodon binding is represented either by a solid circle for Watson-Crick base pairing, or an outlined circle for G:U wobble pairing. (B) Low concentrations of *prfB* cause a longer stall in the recognition of the stop codon, allowing the tRNA-Leu the opportunity to sever the codon:anticodon bonds in the 0-frame and re-pair with UUU in the +1 frame. (C) In the final step, tRNA-Asp reads the GAC codon in the new frame and translation continues.



The pairing of a slippery codon immediately upstream of a second stall-inducing codon is found in almost all +1 translational frameshifts. Secondary structure and stimulating sequences are also common, but are not as ubiquitous. In our investigation of the chelonian frameshift sites, we attempt to identify any programmed frameshift mechanism elements that are present.

Yeast TY Elements

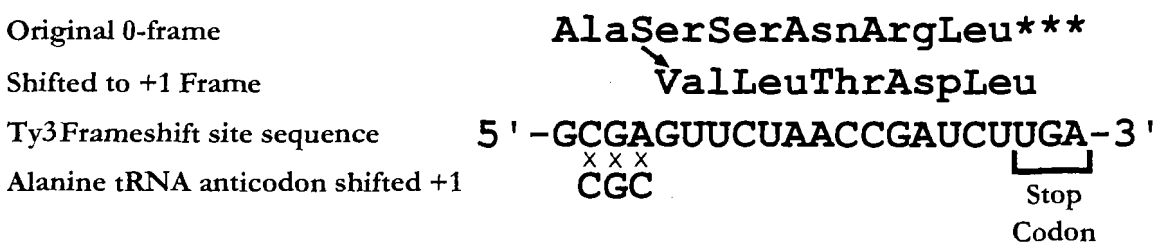
A second example of +1 frameshifting is found in yeast (*Saccharomyces cerevisiae*) Ty1 and Ty3 elements (Belcourt and Farabaugh, 1990, Farabaugh *et al.*, 1993). The Ty elements are retrotransposons containing two genes, *gag* and *pol*. The 3' end of the *pol* gene overlaps the first 38 nucleotides of *gag* in the +1 frame. In these

elements, a *gag-pol* fusion polypeptide, whose production requires a frameshift, is an essential protein. In Ty1, the frameshift site is CUU AGG C, as written in codons of *gag*. The second codon in the frameshifting heptamer, AGG, is rarely used and is again thought to stall the ribosome. Like *prfB*, the peptidyl codon here is a leucine, decoded by the tRNA-Leu (CUN) (anticodon UAG) that is able to re-pair in the +1 frame, moving from CUU in the original reading frame to UUA in the +1 frame, resulting in decoding the sequence as CUU A GGC. The leucine tRNA recognizes the four CUN codons and only weakly pairs with CUU, more readily allowing the disassociation required for shifting and re-pairing.

In Ty3, the sequence surrounding the frameshift is GCG AGU U in *gag*, or read as *gag-pol* fusion codons, GCG A (skipped) GUU (Figure 4). As in Ty1, Ty3 has an in-frame AGN, the rarely used group of serine codons in yeast hypothesized to cause the pause important for frameshifting (Farabaugh *et al.*, 1993). Although no known secondary structure is present, a contextual sequence of 12 to 14 nucleotides immediately downstream affects frameshifting (Farabaugh, 1996).

Figure 4 **Yeast Ty3 element *gag-pol* fusion frameshift site.**

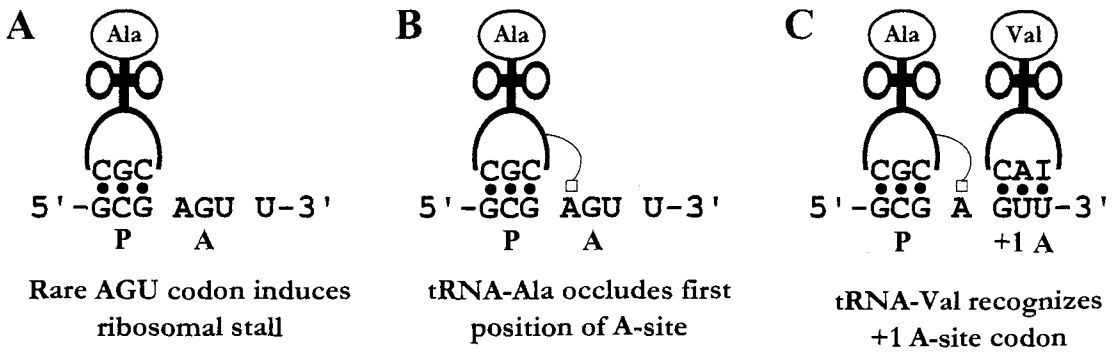
The Ty3 *gag-pol* frameshift site does not appear able to allow the P-site tRNA to shift to the +1 codon. Instead, the P-site tRNA is thought to influence the A-site codon so that the incoming tRNA binds in the +1. The triple mismatch in the between the P-site +1 codon and the anticodon of tRNA-Ala (GCG), the tRNA that would have to shift, is illustrated here.



The mechanism for frameshifting in Ty3 is cause for more controversy than in Ty1, considering that the P-site codon is unable to re-pair well in the +1 frame with tRNA anticodons to facilitate the shift. Instead the peptidyl tRNA is believed to somehow occlude access to the adenosine, the next nucleotide 3', or facilitate +1 binding of the A-site tRNA, either way resulting in the incoming A-site tRNA pairing with the +1 A-site codon (Figure 5). In Ty3, both the peptidyl shift site anticodon and tRNA levels are important factors in frameshifting (Li *et al.*, 2001, Raman *et al.*, 2006). All possible codons were tested in the P-site, and only a few caused observable increases in frameshifting levels, corresponding to eight different tRNAs. Interestingly, along with the original tRNA-Ala (GCG), tRNA-Leu (CUN) was also found to stimulate frameshifting (Vimaladithan and Farabaugh, 1994). It is this tRNA that is used to decode the last 0-frame codon in all known avian and chelonian mitochondrial frameshift sites.

Figure 5 **Proposed Ty3 *gag-pol* frameshift mechanism.**

(A) tRNA-Ala decodes the GCG codon, and moves to the P-site. A pause in translation is then caused by the rare AGU serine codon. (B) The tRNA-Ala is thought to somehow occlude the first position of the 0-frame A-site codon as opposed to re-pairing in the +1 frame. (C) The occlusion of the first position adenosine of the A-site codon allows the GUU valine codon to be read by tRNA-Val, and translation continues in the +1-frame.



Mammalian antizyme

The final +1 programmed translational frameshift example, and the only known mammalian chromosomal frameshift, occurs in ornithine decarboxylase antizyme, a protein with an important role in polyamine biosynthesis. The antizyme protein binds to ornithine decarboxylase and targets it for degradation. Protein sequencing of active antizyme product has revealed that a frameshift occurs at UCC UGA N (Matsufuji *et al.*, 1995). As is evident from the previous examples, if the A-site triplet is not recognized swiftly by a tRNA or release factor, the resulting pause can be an opportunity for the ribosome to shift frames. Here, a delay in recognition of the UGA stop codon by its equivalent release factor produces the translational pause. Also, like the Ty3 element in yeast, in the antizyme frameshift the P-site tRNA is unable to re-pair in the +1 frame. Through mutational analysis, the +1 shift was found to not involve the peptidyl tRNA-Ser (UCN) moving from UCC to CCU (Matsufuji *et al.*, 1995), but instead the first U of the stop codon UGA was found to be blocked in some way. The mechanism that tRNA-Ser (UCN) uses for this remains unclear. Several other tRNAs have been shown to have a similar ability, but when used in the same sequence, they produce much lower levels of frameshifting (Matsufuji *et al.*, 1995), suggesting there are other signals involved that remain undiscovered.

The antizyme frameshift site contains some other elements that seem at odds with most proposed frameshift mechanisms. The first is a 3' pseudoknot that is 3 nucleotides from the frameshift site shown to increase frameshifting from 2.5- to 5-fold, as revealed through deletion mutations to the pseudoknot sequence. This is the only known example

of a +1 frameshift stimulated by a pseudoknot. A second uncommon feature is a 5' sequence of 50 nucleotides that stimulates frameshifting two-fold (Matsufuji *et al.*, 1995).

These three well-studied examples of programmed translational frameshifting have a few primary similarities that appear to be generic of all +1 shifts. All three employ a rare or nonsense codon immediately after the last in-frame codon. In order to frameshift at the required efficiency, all genes code for the use of a frameshift-capable tRNA at the P-site codon. In bacterial *prfB* and yeast Ty1, it is a tRNA that is able to slip easily and re-pair well in the +1 frame. In yeast Ty3 and mammalian antizyme, it appears that there is some other property of the P-site tRNA that increases frameshifting levels by causing the incoming A-site tRNA to bind +1. Other secondary features such as the various frameshift stimulating RNA sequences or structures seen in the above examples are not as well-conserved, though they appear to be often present in one form or another.

Frameshifting in animal mitochondrial genomes

Discovery of a frameshift mutation in the *nad3* gene of birds and turtle

A mitochondrial frameshift mutation was discovered in the NADH dehydrogenase subunit 3 (*nad3*) gene in the ostrich (*Struthio camelus*) by Härlid *et al.* in 1997. At the time, they hypothesized that the mutation caused *nad3* to be terminated prematurely. Later, Mindell *et al.* (1998) examined the same site in a number of other birds, a turtle, and other vertebrates. They found that the inserted extra nucleotide observed in the ostrich *nad3* gene was present in many other birds as well as the painted turtle (*Chrysemys picta*). Moreover, if it were removed, a full length *nad3* open reading frame would result. They hypothesized that the *nad3* gene was not truncated in the

ostrich or in the other vertebrates where the extra nucleotide was present, but that it was skipped to produce a full length protein product.

In the chicken (*Gallus gallus*), the *nad3* gene is 354 base pairs (117 amino acids) in length (Desjardins and Morais, 1990). If translated with standard decoding, the ribosome encounters a stop codon shortly after the extra nucleotide and as a result is only 207 base pairs (68 amino acids) long. It is worth noting that in Desjardins and Morais' original description of the chicken mitochondrial genome, they did not report an extra nucleotide in *nad3*, though it is present in all sequences of this region from subsequent studies. Mindell *et al.* (1998) offer a number of arguments that suggest that the extra nucleotide is skipped and the gene translated in its entirety. They propose that if the gene had a stop codon somewhere internally, the remaining DNA 3' of this site would have all selective constraints relaxed, something that would be evident in phylogenetic comparisons. They point out that removing the extra nucleotide restores the reading frame and all the downstream sequences from that point on show a high level of sequence conservation. After comparing the two groups of mitochondrial genomes, those with and those without the extra base, it appears that the gene remains functional in its entirety in both groups. Non-synonymous mutations in this gene have been shown to have lethal effects in various vertebrates (*e.g.* MacFarland *et al.*, 2003), so it is unlikely that the gene translates as a truncated protein. Further, in an investigation into mitochondrial gene copies within the chicken nuclear genome, no full or partial copies of *nad3* were discovered (Pereira and Baker, 2004). These results support the conclusion that the *nad3* gene remains functional in the species with the extra nucleotide, and that the nucleotide is

either removed through an editing mechanism or is skipped during translation by translational frameshifting.

+1 Frameshift in a genus of ants and the eastern oyster

Along with the birds and painted turtle found in the Mindell study, more examples of frameshifting in mitochondria have been found in a variety of different taxa recently. In many species of ants in the genus *Polyrhachis*, frameshift insertions have been found in the cytochrome oxidase b gene (*cytb*). Of thirty species studied, 12 have one or two +1 frameshift sites, with 4 different sites affected (Beckenbach *et al.*, 2005). Most of these frameshifts appears to employ a mechanism similar to that found in yeast Ty1 elements, where a rare codon is present in-frame at a site immediately downstream of a slippery codon. The resulting pause allows the tRNA at the P-site to shift +1 where it has a near cognate binding to this new codon (Beckenbach *et al.*, 2005). One species with the most common *Polyrhachis* frameshift site that is likely decoded by this mechanism also has another frameshift insertion that does not have good +1 pairing for the P-site tRNA. This may suggest that both proposed mechanisms of +1 translational shifting, P-site shifting of the t-RNA to the +1 frame and occlusion (or a related mechanism) of the first position of the A-site codon, can be present in the same translational system.

A final mitochondrial frameshifting example is found in the Eastern Oyster, *Crassostrea virginica* (Milbury and Gaffney, 2005). As in ants, the frameshift insertion was found in the *cytb* gene, only one nucleotide away from the most common site in the *Polyrhachis* study. The frameshift occurs over TAC T AGG, shown as codons of the conserved reading frame. In another oyster species, this site is TAT AGG, suggestive of a

C insertion into the last position of the tyrosine codon. In this case, the translational pause occurs at a TAG stop codon, and the authors suggest that this is another example of occlusion of the first position of the A-site codon. Figure 6 shows a summary of the known animal mitochondrial frameshift sites.

All of the previous examples were provided in an effort to detail what is known about translational frameshifting, with a focus on +1 shifting, the only type observed in mitochondria to date. The amount of evidence for mitochondrial frameshifts continues to grow, allowing more detailed analyses of what is required for a mitochondrial programmed translational frameshift to produce functional proteins.

Mitochondrial genome as a model system for recoding of translation

The mitochondrial genome in animals is typically 16,000 to 20,000 base pairs in size, inherited maternally. Contained within it are 13 protein coding genes, 2 ribosomal RNAs, and 22 transfer RNAs. It is these 22 tRNAs that are responsible for the translation of the 13 mitochondrially-encoded proteins. Mitochondria have evidently evolved to reduce genome size, and as a result use this minimal number of tRNAs to decode a far greater number of sense codons. With 22 tRNAs decoding 59-62 sense codons, about two thirds of the time codons are decoded with non-cognate tRNA anticodon:codon pairings (Figure 7). Also of note is a genetic code in mitochondria that differs from the standard code, especially with regards to the AGR group of codons (AGA and AGG). While they code for arginine in the standard code, in vertebrate mitochondria AGR codons are thought to be termination codons, though they are not often used.

Figure 6 **Summary of the known mitochondrially-encoded programmed translational frameshift sites in animals.**

Examples of all known animal mitochondrial frameshifts are listed. Codons are spaced in both the original and shifted frames. The single nucleotide of the Nucleotide Skipped column is not translated due to the ribosomal frameshift. In all cases, frameshifting allows production of a protein with conserved amino acid sequence. Note that the single nucleotide is not necessarily the one that was inserted in the original frameshift mutation.

Organism	Gene	Original Frame	Nucleotide Skipped
Ants			
P. sexspinosa	*cytb*	GGT AGT AAC G S N	GGT A GTA ACC G V T
	cytb	GGG AGC AAC G S N	GGG A GCA ACC G A T
P. phryne	*cytb*	TGG AGT ATG W S M	TGG A GTA TGG W V W
	cytb	ATA GGT AAC M G N	ATA G GTA ACC M V T
Birds			
S. camelus	*nad3*	CTC AGT AGC L S S	CTC A GTA GCA L V A
G. gallus	*nad3*	CTC AGT AGC L S S	CTC A GTA GCC L V A
Turtles			
C. picta	*nad3*	CTG AGT AGC L S S	CTG A GTA GCA L V A
P. subrufa	*nad3*	CTT AGA ACC L * T	CTT A GAA CCA L E P
	nad4l	CTT AGA AGG L * *	CTT A GAA GGC L E G
	nad4l	CTT AGA TAT L * Y	CTT A GAT ATA L D M
Oyster			
C. virginica	*cytb*	TAT TAG GGG Y * G	TAT T AGG GGC Y S G

Figure 7 **Vertebrate mitochondrial genetic code and differences from the standard code.**

With the exception of serine and leucine residues, each individual amino acid is represented by a single, mitochondrially-encoded tRNA. The complete mitochondrial genetic code is listed, along with the differences between it and the standard code. Of note are the various changes to the stop codons between the two codes. The AGR codons, AGA and AGG, are terminators in the vertebrate mitochondrial code, instead of coding for arginine as they do in the standard code. The UGA terminator has also been redefined as tryptophan in vertebrate mitochondria.

Vertebrate Mitochondrial Genetic Code

TTT F Phe	TCT S Ser	TAT Y Tyr	TGT C Cys
TTC F Phe	TCC S Ser	TAC Y Tyr	TGC C Cys
TTA L Leu	TCA S Ser	TAA * Ter	TGA W Trp
TTG L Leu	TCG S Ser	TAG * Ter	TGG W Trp
CTT L Leu	CCT P Pro	CAT H His	CGT R Arg
CTC L Leu	CCC P Pro	CAC H His	CGC R Arg
CTA L Leu	CCA P Pro	CAA Q Gln	CGA R Arg
CTG L Leu	CCG P Pro	CAG Q Gln	CGG R Arg
ATT I Ile i	ACT T Thr	AAT N Asn	AGT S Ser
ATC I Ile i	ACC T Thr	AAC N Asn	AGC S Ser
ATA M Met i	ACA T Thr	AAA K Lys	AGA * Ter
ATG M Met i	ACG T Thr	AAG K Lys	AGG * Ter
GTT V Val	GCT A Ala	GAT D Asp	GGT G Gly
GTC V Val	GCC A Ala	GAC D Asp	GGC G Gly
GTA V Val	GCA A Ala	GAA E Glu	GGA G Gly
GTG V Val i	GCG A Ala	GAG E Glu	GGG G Gly

Differences between the Vertebrate Mitochondrial Genetic Code and the Standard Code

	Vertebrate Mitochondrial		Standard	
AGA	Ter	*	Arg	R
AGG	Ter	*	Arg	R
AUA	Met	M	Ile	I
UGA	Trp	W	Ter	*

Purpose of this study

To investigate frameshifts in essential mitochondrial genes, we began with an organism that was previously reported to carry a frameshift mutation. Initially, we wanted to confirm the presence of the extra nucleotide in the mitochondrial DNA and to distinguish between RNA editing or translational level compensation for the frameshift mutation. Using DNA samples from the domestic chicken (*Gallus gallus*), shown by Mindell *et al.* to carry an extra nucleotide in its mitochondrial *nad3*, we sequenced a region of the mRNA transcript of this gene, and compared that to the mitochondrial DNA sequence. We were able to confirm the presence of the frameshift in both the mitochondrial DNA and the mature mRNA transcript.

Mindell *et al.* (1998) found that approximately two-thirds of all birds showed the frameshift mutation. They also found that the painted turtle (*C. picta*) had the extra nucleotide. The question that remained is whether the mutation was widespread in turtles as it is in birds. To answer this, we carried out a survey of *nad3* sequences from a range of different turtles. This not only allowed us to analyze turtles for the presence of the frameshift, but also enabled sequence comparisons with birds containing the frameshift. Ideally, these comparisons would reveal specific conserved elements intrinsic in stimulating +1 translational frameshifting.

Examination of the complete turtle mitochondrial genome sequences in the GenBank database has led to two interesting discoveries. The first is the presence of three frameshift anomalies within the mitochondrial genome of the African sideneck turtle, *Pelomedusa subrufa* (NC_001947, Zardoya and Meyer, 1998). This observation is

particularly intriguing as all three sites appear to be unique to this species, unlike the common *nad3* mutation insertion. The second discovery is not a feature of frameshift sites, but an important qualification. For some other GenBank database entries, the sequences are evidently not correct – in particular, for some it appears that sequences were modified in an unfortunate attempt to force them to conform to preconceived expectations. Zardoya and Meyer (1998) were thorough in their analysis of the frameshift regions in the African sideneck turtle, and carefully documented the presence of nucleotides at all sites, even those that appeared to disrupt the conserved reading frame. Therefore, we feel comfortable using these sequences in the analysis of frameshifts in mitochondria, while other GenBank complete mitochondrial genome sequences, particularly those submitted prior to universal recognition of the common *nad3* frameshift site, remain somewhat suspect and are not included in comparative analyses.

We hypothesize that with respect to mitochondrial translation systems, organisms that are able to tolerate frameshift mutations at one site in their genome are more able to tolerate similar mutations elsewhere, as appears to be the case in *Polyrhachis* ants as well as in *P. subrufa*. This requires correct sequences, free of any post-sequencing adjustments to make them consistent with an impression of what the genome should contain. The decision to sequence the complete red-eared turtle (*Trachemys scripta*) mitochondrial genome was made partially for this reason, allowing accurate analysis of the known frameshift region and a check for others. Our hypothesis suggests that species that are able to tolerate the common *nad3* frameshift mutation may have frameshift insertions elsewhere, so initially I screened the genome for the presence of the extra nucleotide

within *nad3* by sequencing a small region around the site. Once this was confirmed, the complete genome was sequenced.

A note on the challenges of recoding research

Even today, with many published examples of a variety of recoding events, when such events are discovered in sequences, they are often dismissed as sequencing errors. Sequencing techniques and lab techniques exacerbate this, with the focus on some other feature of the genetic material. Often these events are discarded as trivial, or worse, "repaired". With many more recognized examples of programmed translational frameshifting and an increased reliance on automated sequencing and short read, high throughput machines, researchers will hopefully become more aware and thorough in documenting these interesting genetic events.

MATERIALS AND METHODS

nad3 DNA and mRNA in chickens

Chicken RNA and DNA samples

Three fresh chicken livers were obtained from Lilydale Poultry in Port Coquitlam, British Columbia, Canada, on Friday, November 12, 2004. All RNA extractions were done the same day. Dr. James Stewart provided valuable assistance with the RNA extractions.

Crude mitochondria isolated from fresh chicken liver

Small segments of the chicken liver were removed with a sterile scalpel and placed into an ice-chilled Wheaton homogenizer along with 250 ul of cold MSB buffer (210 mM mannitol, 70 mM sucrose, 50 mM Tris-Cl pH 7.5, 10 mM EDTA) per gram of tissue. Disruption of the tissues was carried out with ten strokes of the homogenizer. Mitochondrial isolation was conducted at cold temperatures, either with the tubes chilled on ice or in a 4 °C cold room.

After transferring 1.0 ml divisions of aqueous homogenate into 1.5 ml Eppendorf tubes, the tubes were centrifuged for 5 minutes at 4000 rpm at 4 °C to remove cellular detritus. The resulting supernatant was moved into 1.5 ml Beckmann ultracentrifuge tubes and centrifuged for 20 minutes at 20,000 rpm in a Beckmann TLA-45 centrifuge rotor to pellet the mitochondria. The pellet was then resuspended in 1.0 ml of MSB and

centrifuged again at 20,000 for 20 minutes. RNA and DNA extraction were performed on the pellet immediately once it had been resuspended in 100 ul of 1X TE (10mM Tris-Cl, 1 mM EDTA, pH 7.4).

DNA extraction

DNA extraction followed a protocol similar to Stewart, 2005. After placing each of two tissue samples in Beckman ultrafuge 1.5 ml Eppendorf tubes, 200 ul of proteinase K buffer (0.01M Tris at pH7.8, 0.005 M EDTA, 0.5% SDS, 50 ng/ul proteinase K) was added and the tubes were incubated for one hour at 60 °C. Following the digestion, the solution was extracted using an equal volume of Tris-buffer saturated phenol (pH 7.6). The aqueous layer was removed and then 1 volume of SEVAG solution (chloroform:isoamyl alcohol, 24:1) was added. Again the aqueous layer was removed, and 2.5 volumes of cold 95% ethanol were added. The tubes were vortexed by hand and put in a -20 °C freezer for overnight precipitation. The following day, the tubes were centrifuged for 20 minutes at 15,000 rpm to precipitate the DNA. The pellet was washed with cold 70% ethanol and centrifuged again. After this was repeated twice, the DNA was resuspended in 100 ul of ddH_2O.

RNA extraction

RNA extraction requires special precautions to prevent the contamination of samples with any RNases, which are omnipresent in our environment. To ensure this, all experiments and manipulations with RNA were done on a bench top devoted solely to this purpose. This included dedicated pipettors and pipette tips certified to be RNase free.

Prior to any experiments, all equipment and surfaces were treated with RNase Erase (ICN Biomedicals Inc.). All primers were dissolved and diluted with DEPC-treated H_20.

For RNA isolation, the Ambion® Inc. ToTALLY RNA™ RNA isolation kit was used, following the protocol in the kit's documentation. Lysis of the crude mitochondrial pellet was accomplished through the addition of the supplied Denaturation Solution and one minute of vortexing. Disrupted mitochondria were extracted with 150 ul of the supplied Phenol:Chloroform:Isoamyl Alcohol solution, vortexed for one minute, and put on ice for 5 minutes. After centrifuging for 5 minutes at full speed in an Eppendorf 5415C desktop centrifuge the aqueous phase was removed through careful pipetting. To this, one tenth of its volume of the kit's sodium acetate solution was added. To extract the solution, 150 ul of the provided Acid-Phenol:Chloroform solution was added, vortexed for one minute, and put one ice for 5 minutes. After centrifugation, the aqueous phase was moved to an RNase-free 0.6 ml Eppendorf tube together with an equal volume of isopropanol. This was left overnight in a -20 °C freezer to precipitate.

Final recovery of the RNA involved centrifuging precipitated RNA at full speed in an Eppendorf 5415C desktop centrifuge for 20 minutes. The liquid was removed, and the pellet washed with 200 ul of 70% ethanol. After another round of centrifugation, the pellet was washed a second time and then allowed to dry near a Bunsen burner for 20 minutes. One tube was placed in a -80 °C freezer dry for long term storage, and 100 ul of DEPC treated H_2O with 0.1 M EDTA supplied with the kit was used to dissolve the pellet in the second for immediate use. Contaminating DNA was removed using Ambion® TURBO™ DNase (RNase-free). The 10X DNase buffer was added to the RNA samples, along with units (1 ul) of TURBO™ DNase for each 49 ul of RNA-DNase buffer solution.

The solution was incubated at 37 °C for 30 minutes. Extraction was done as described previously, but with only one-tenth of the volume of denaturation solution. Two extraction products were produced by this procedure.

Reverse-transcription PCR

The Enhanced Avian HS RT-PCR Kit (Sigma®) was used for Reverse-Transcriptase PCR (RT-PCR) reactions. All reactions were done in 25 ul total final volumes. The RT-PCR reactions were all prepared in RNase-free certified 0.2 ml thin-walled PCR tubes. In each reaction, the final concentrations were 200 uM of each dNTP, 3.0 mM MgCl2, 0.4 uM of each primer, 0.4 units/ul of RNase Inhibitor Enzyme (supplied), 0.4 units/ul of eAV-RT reverse transcriptase, and 0.05 units/ul of Jumpstart AccuTaq™ LA DNA polymerase. These were all in a 1X reaction buffer, with about 0.4 ng (added as 1 ul) of RNA extract in each reaction. The primers used are listed in the Appendix.

The RT-PCR thermocycler protocol began with a 60 minute incubation step at 42 °C to enable reverse transcription. This was followed with 2 minutes at 94 °C, and then 35 cycles of 94 °C for 15 seconds, 55 °C for 30 seconds, and 68 °C for two minutes. Finally, a 5 minute extension step at 68 °C concluded the reaction.

Control reactions to test RNA samples for DNA contamination

Along side all RT-PCR reactions, control reactions were run to ensure that the total RNA extract does not have contaminating DNA present. This is especially important here as we are investigating if the extra nucleotide present in the DNA is transcribed to RNA and not edited out. Two tubes were set up to test each of the two extractions. Prior

to the addition of the RT-PCR reaction mixture, each tube received 1 ul of RNase and 1 ul of ddH_20 (to bring the final reaction volume equal to 25 ul). The A and B samples were added to their respective reactions.

DNA sequencing

All sequencing was done by the University of Calgary's Core DNA & Protein Services. Samples were sent to the University of Calgary premixed with both primer and template present. In each 12 ul sample sent, there was 3.2 umol of primer, and 100 ng per kb length of template. Core DNA & Protein Services uses 3730 Genetic Analyzer ABI DNA Sequencing Instrument as their primary sequencer with the BigDye Version 3.1 sequencing kit. Once the samples are sequenced, a pdf of all sequence reactions, AB1 trace files, and sequence text files are posted to a central server where they were then downloaded locally.

Complete genome sequencing

Specimen collection

A skin sample from a red-eared slider turtle (*Trachemys scripta*) was obtained from the Reptile Refuge in Surrey, British Columbia, Canada, on November 15, 2006. The sample was approximately one inch square, and was collected and identified by Paul Springate of the Refuge. The tissue was then placed in a sterile container and allowed to completely dry out prior to DNA extraction. Prior to DNA extraction, the tissue was cleaned with 70% ethanol and allowed to completely dry overnight.

DNA extraction

DNA extraction was done identically for the all tissue samples as described previously for chicken liver, after dividing the skin sample into four smaller pieces and digesting two separately. The remaining two were stored in sterile containers for archival purposes.

PCR amplification

The mitochondrial genome was amplified in overlapping fragments using the primer pairs listed in the Appendix. Heterologous primers were designed using an alignment of six turtle and tortoise sequences available in the GenBank database (accession numbers: NC_000886, NC_001947, NC_002073, NC_002780, NC_006082, NC_006132). Once portions of the genome were sequenced, sequence specific primers were designed to amplify remaining sections. For all PCR amplifications, either 5X or 10X dilutions of the original DNA extract were used. Both an Eppendorf Mastercycler gradient thermocycler and an Eppendorf Mastercycler personal were employed in PCR reactions. All PCR reactions used TaqPro™ from Danville Scientific in 25 ul volumes. The TaqPro™ buffer was diluted down from the supplied 10X to 1X, and appropriate volumes of each of the following were added to get final concentrations of 0.2 nM of each dNTP, 400mM of each primer, 2.0 mM MgCl, and 0.5 U of TaqPro™ polymerase in each 25 ml reaction.

PCR cycling began with a 1:30 minute denaturation at 94, and then continued with four cycles of 20 seconds of denaturation at 93 °C, 30 seconds of annealing at 45 °C, and extension for 30 seconds at 72 °C. Following the initial 4 cycles, 35 cycles were done

with the only difference being an anneal temperature of 50 °C instead of 45 °C. Some primer pairs produced non-specific results, and required that the anneal temperature be raised. In these cases, the anneal temperature was raised to 52 degrees for all cycles, with 35 total cycles. All other temperatures and times were kept the same.

The results of PCR reactions were investigated using 0.5% agarose gels with a buffer of 0.5X TAE (20 mM Tris-HCl, 5 mM EDTA, 10 mM acetic acid). One ul of loading buffer was added to 4 ml of reaction solution and was run alongside 200 ng of 1 Kb Plus DNA Ladder from Invitrogen® Life Technologies.

PCR product purification

Amplified products that produced single bands of approximately the correct estimated size were cleansed using the QIAquick™ PCR Purification kit from QIAGEN® following the steps outlined in the documentation. This is done to remove all extra PCR amplification by-products (extra nucleotides, primers, MgCl) in order to prepare the product for sequencing. The process begins by mixing the original PCR reaction with five volumes of the supplied Buffer PB, a solution containing chaotropic salts that disrupts non-bonding forces, used in the binding of the PCR product to the supplied column upon centrifugation. The bound product was then washed with 750 ul of another proprietary solution, containing ethanol, via centrifugation. After another blank centrifugation step to remove all remaining traces of the ethanol solution, the column was dried thoroughly for at least three hours. Adding 30 ul of ddH_20 and centrifuging eluted the DNA into a 1.5 mL Eppendorf tube.

Sequencing

All sequencing was conducted at the University Core DNA and Protein Services at the University of Calgary as described previously.

Sequence assembly

Genome assembly was done manually using BioEdit (version 7.0.4.1, February 13, 2005, Hall, 1999) from the sequence and trace files generated by the automated sequencing. tRNAScan-SE (Lowe and Eddy, 1997) was used to do tRNA searches using only organellar tRNAs, the vertebrate mitochondrial genetic code, and a COVE score of 5 as the constraints.

nad3 region analysis

Tissue samples

DNA was extracted from various tissue sources. DNA samples were obtained from blood from *Pelomedusa subrufa* (African side-neck turtle), *Rhinoclemmys pulcherrima manni* (Central American wood turtle), *Batrachemys* (formerly *Phrynops*) *nasuta* (Toad-headed turtle), *Staurotypus triporcatus* (Mexican giant musk turtle), *Macrochelodina* (formerly *Chelodina*) *parkeri* (Parker's side-necked turtle), *Mauremys* (formerly *Annamemys*) *annamensis* (Annam leaf turtle), *Geochelone carbonaria* (Red-footed tortoise), *Indotestudo forsteni* (Elongated tortoise); scutes from *Alligator mississippiensis* (American alligator); a tooth from *Caiman crocodiles* (Spectacled caiman); shell and carapace pieces from *Sternotherus odoratus* (Musk turtle), *Chelydra serpentina serpentina* (Snapping turtle), *Terrapene carolina carolina* (Box turtle),

Apalone ferox (Florida softshell turtle); and shell clippings from *Malaclemys terrapin* (Diamondback terrapin), *Macroclemys temminckii* (Alligator snapping turtle), *Kinosternon subrubrum* (Common mud turtle), *Geochelone sulcata* (Spur-thighed tortoise), and *Clemmys guttata* (Spotted turtle). DNA was also extracted from a skin sample from *Trachemys scripta* (Red-eared slider) and dried body segments from *Tantilla coronata* (Crown snake). Samples of *A. mississippiensis, S. odoratus, C. serpentina serpentine, T. carolina carolina, A. ferox,* and *T. coronata* were collected by Dr. A.T. Beckenbach between 1990 and 2000 from various locations in the continental United States. All blood samples were obtained courtesy of Eric Holt of Empire of the Turtle in Yalaha, Florida. Shell clippings, skin samples, and the caiman tooth were all collected at the Reptile Refuge in Surrey, British Columbia with the valuable assistance of Paul Springate.

Initially, powder from the turtle shells was used for DNA extraction. This powder was mechanically ground off the sample using an acid treated rotary file. Acid treating consisted of a submersing the file in dilute HCl and then neutralizing the acid with NaOH, and finally a prolonged submersion in double distilled water. Though this method did produce DNA that was adequate for the needs of this study, due to concerns about contamination and the possibility of trace amounts of acid in the file reducing the quality of the DNA samples the technique was abandoned in favour of a liquid nitrogen protocol. Carapace, tooth and scute samples were ground to a powder using an acid-treated mortar and pestle in the presence of liquid nitrogen, and the resulting powder was used for DNA extraction.

DNA extraction

DNA from tissue samples was extracted through a phenol-chloroform protocol. Approximately two volumes of proteinase K buffer (0.01M Tris at pH7.8, 0.005 M EDTA, 0.5% SDS) were added to the powder from all skin, carapace, and scute samples in 1.5 ml Beckman ultracentrifuge tubes. After adding 50 ng/ul of proteinase K, the resulting solution was incubated at 60°C for at least one hour. As the consistency of the powdered sample varied, a further volume of proteinase K buffer solution was added if the mixture appeared too viscous.

DNA from blood samples was extracted using variation of the phenol-chloroform protocol used for bone and skin. The blood was extracted in Yalaha, Florida and placed on in sodium heparinized tubes. Forty-eight hours after they were drawn, the samples were received and DNA was immediately extracted.

PCR amplification

PCR amplification was performed with an Eppendorf Mastercycler gradient and personal thermocyclers, with protocols similar to those used in the PCR amplifications of the complete *T. scripta* genome. For PCR amplification of DNA, both Quiagen's® Taq DNA polymerase and Danville Scientific's TaqPro™ DNA polymerase enzymes were used. PCR products were run out on 0.5% agarose gel, with reactions that produced clean bands prepared directly for sequencing and reactions with multiple bands gel isolated prior to sequencing.

Sequencing

All sequencing was conducted at the University Core DNA and Protein Services at the University of Calgary as described previously.

Bioinformatics

Secondary structure examinations

ClustalW (Thompson *et al.*, 1991) was used to prepare alignments that were then analyzed with the Alifold web server that is based on the Vienna RNA Secondary Structure Prediction package (Hofacker *et al.*, 2002). The Alifold program predicts a consensus RNA secondary structure from a set of aligned sequences. Various comparison groups were used in this analysis (Table 2). Default values were used in all settings. The MARNA (Multiple Alignment of RNA) web server (Siebert and Backofen, 2005) was also used to predict possible secondary structures for the groups used with Alifold. Once again, default values were used in all settings. Finally, the mFOLD web server (Zuker, 2003) was used with default values to calculate possible RNA secondary structure around each of the two frameshift sites found within the *T. scripta* mitochondrial genome. Both a small region of the surrounding sequence of the frameshift site 68 nucleotides long and the complete gene were analyzed in this manner.

Table 2 **Frameshift site mRNA secondary structure analysis.**

Groupings used for consensus sequence RNA secondary structure prediction. In order to determine what features are essential for frameshifting, sequences were divided into two groups, those with and those without the extra nucleotide, and then further divided as described below. Wherever GenBank sequences were used, this included all organisms with complete mitochondrial genomes in the database. These groups were used in both Alifold and MARNA. Note that comparisons using the GenBank database did not use *P. sinensis*, as its highly irregular sequence is inconsistent with any other chelonian and may be erroneous.

nad3 Small Region	Complete *nad3* Gene	Complete *nad4l* Gene
Sequenced Turtles	GenBank Turtles + *T. scripta*	GenBank Turtles + *T. scripta*
Genbank Turtles (*Testudines*)	GenBank Birds (*Aves*)	GenBank Birds (*Aves*)
Turtles Combined	All combined	All combined
GenBank birds (*Aves*)		
All combined		

Codon usage

Overall codon usage and relative synonymous codon usage (RSCU) were calculated for the 64 possible codons used in the *T. scripta* mitochondria. Overall codon usage was calculated by dividing the number of times a particular codon was used by the total number of codons in the mitochondrial genome. RSCU was calculated for each codon by dividing the total number of times a particular codon was used by the product of the number of codons in the amino acid codon family and the total number of codons that code for that particular amino acid. For comparison, this number was also normalized by multiplying the number of times the codon is used by 64 before dividing by the total number of codons in the genome.

RESULTS

The chicken *nad3* frameshift site

To confirm the presence of the extra nucleotide reported by Mindell *et al.* (1998) and to determine whether it is removed by some process of RNA editing, a small region of the *Gallus gallus* mitochondrial genome around the *nad3* frameshift site was sequenced along with a corresponding region of the polyadenylated *nad3* mRNA transcript. Both sequences show the presence of the extra frameshift-causing nucleotide at position 174 in the *nad3* gene (Figure 8). The two sequences also align perfectly with the sequenced chicken from Mindell *et al.* (1998). This result appears to eliminate RNA editing as a possible mechanism for accurate *nad3* translation and suggests that the frameshift is compensated for by a translational mechanism allowing it to be read through. To allow for the production of a functional *nad3* polypeptide, the ribosome somehow must be instructed to shift frames at this particular site and continue translation in the correct +1 frame.

Figure 8 **Chicken mitochondrial genomic sequence and mRNA sequence of *nad3* frameshift region.**

Our two sequences, from the mitochondrial genomic copy of *nad3* and the corresponding mature mRNA transcript display complete conservation to the sequence obtained by Mindell *et al.*, including retention of the inserted nucleotide within the mRNA transcript. In Desjardins and Morais' original sequence of the chicken mitochondrial genome, they did not report the extra nucleotide. The mRNA was amplified with an internal primer paired with a poly-T primer, to ensure mature mRNA sequence was obtained.

Desjardins and Morais (NC_001323)	5'-ATCCGATTCTTCCT-AGTAGCCATCCTATTCCTTTT-3'
Mindell *et al.* (AF076356)	5'-ATCCGATTCTTCCTCAGTAGCCATCCTATTCCTTTT-3'
mtDNA, this study	5'-ATCCGATTCTTCCTCAGTAGCCATCCTATTCCTTTT-3'
mRNA, this study	5'-ATCCGATTCTTCCTCAGTAGCCATCCTATTCCTTTT-3'

Do these frameshifted genes remain functional?

We have shown the extra nucleotide is not removed prior to translation in chicken mitochondria. The question that remains is whether or not genes with frameshift mutations remain functional. The case of *nad3* in the chicken was outlined above. The absence of any close relative of the essential *nad3* gene in the nuclear genome, and strong conservation of sequence downstream of the frameshift site are strong arguments in favour of the mitochondrial *nad3* gene retaining its function. Indeed, the principal evidence that genes remain functional even with a frameshift-causing nucleotide is that the sequence is conserved in the +1 reading frame after the insertion. More specifically, substitutions in frameshifted genes across diverse taxa retain codon positional bias, with

substitutions in the +1 frame after the extra nucleotide concentrated in the 3rd position (Beckenbach *et al.*, 2005). The case is similar in all turtles where the extra nucleotide is present, which strongly suggests that the genes are translated into working proteins.

Characteristics of the *nad3* gene and discovery of the frameshift site

The mitochondrial NADH ubiquinone oxidoreductase chain 3 (*nad3*) gene functions in the respiratory pathway, and is one of seven subunits of the NADH-dehydrogenase complex encoded in the mitochondrial genome (*nad1-6* and *nad4l*). This complex is responsible for the catalysis of the first step of the respiratory electron transport chain, and apparently functional copies of the NADH genes are present in all sequenced vertebrate mitochondria to date. The *nad3* gene in bird and turtle mitochondria is typically 350 nucleotides or 174 amino acids long. Little is known about its longevity, degradation, and amount present. This information could be important with regards to how much protein product is needed, and how often it has to be made as frameshifting could reduce the capacity for protein production.

The extra nucleotide in *nad3* was first revealed in the ostrich (*Struthio camelus*) by Härlid *et al.* in 1997. They hypothesized that the extra nucleotide resulted in *nad3* being prematurely terminated. It was subsequently been found in 44 birds and the painted turtle, and absent in another 15 birds (Mindell *et al.*, 1998). Many other examples have arisen in the time since, with the nucleotide insertion documented in the complete mitochondrial sequences of 14 turtles and tortoises and 37 birds in GenBank. It is absent in all other sequenced reptiles. The *nad3* gene is an essential gene in animal mitochondria, and functional translated proteins are required in all organisms. The

ribosome therefore must have a relatively efficient way of translating over the frameshift disruption caused by the extra nucleotide. Genetic sequence near the frameshift site provides some clues as to how this is accomplished.

nad3 frameshift region features in turtles

We wished to investigate if there were any particular sequences or other features that are conserved in turtles having the frameshift nucleotide that may have a role in frameshift stimulation. This approach is especially powerful if we can subsequently show the absence of these elements in mitochondrial genomes without the extra nucleotide. To do this, we sequenced the region surrounding the frameshift site within the *nad3* gene in 21 different turtles, tortoises, and other reptiles (Figure 9). The extra frameshifting nucleotide was present in 14 of these sequences, all chelonians. Within these taxa, all but the musk turtle (*Sternotherus odoratus)*, Mexican giant musk (*Staurotypus triporcatus)*, and the toad-headed turtle (*Batrachemys nasuta*) showed the extra nucleotide. I also confirmed the presence of a different *nad3* frameshift site upstream from the common site in the African sideneck turtle (*Pelomedusa subrufa)*, as first reported by Zardoya and Meyer in 1998. None of the other reptiles investigated had any frameshift insertion mutations within their *nad3* genes. Close examination of the Parker's sideneck turtle (*Macrochelodina* (formerly *Chelodina*) *parkeri*) reveals one final twist. At the site of what is a highly conserved arginine residue in all other sequenced chelonians, the Parker's sideneck has an AGA codon, a termination codon by the standard vertebrate mitochondrial code.

Figure 9 ***nad3*** **frameshift region for reptiles sequenced in this study.**

Complete list of all sequenced reptiles for the *nad3* gene from positions 132 to 199 relative to the *T. scripta nad3* gene. Numerical key is provided for reference only. The most common insertion appears to be a C, present in 8 of the 13 sequences that show the extra nucleotide, with G and T appearing 2 and 3 times, respectively. The * indicates the location of the nucleotide that must be skipped in order to maintain the conserved reading frame. Nucleotides 44-50 immediately downstream of this site are highly conserved in all organisms with the frameshift-causing nucleotide. This is not the case in organisms lacking the insertion mutation.

```
                          10        20        30        40  *     50        60
Consensus            A-CTAGAATCAGCTCGCCTACCATTCTCAATCCGATTCTTCCTCAGTAGCAATCTTATTCCTCCTATTT
Translation             L  E  S  A  R  L  P  F  S  I  R  F  F  L  V  A  I  L  F  L  L  F

Apalone              .-T..A..A.CATA.....................C.................TC.....T.A..G..C
Chelydra             .-T............AT..........C.............................T.........
Geochelone carbonaria .-........T..C..T.....G.................T.........C...C.......A.....C
G. sulcata           .-...A....T.....T.............T.............................A......
Indotestudo          .-A..A....T.....................................................T.........
Macroclemys          .-T............A.....................T.....................T.........
Rhinoclemmys         .-T..A..A.TA......C.........T........T.........................AT.....
Terrapene            .-...........C.....T.....................T.....................
Malaclemys           .-T.........................C..............G...................T.....
Trachemys            .-T..........C..A..G........C.....G........G...................T.....
Clemmys              .-...........C.....T.....................T.....................
Macrochelodina       C-..T......A..T.A..G........T..TA..........T.........C..........T...
Mauremys             .-...A..A.CA.C..T.C............T........T..T.........T........A.....C

Pelomedusa           .CT.....C.....T.AT...........G.T.....T.....T-A.C...........T.........
Staurotypus          .-T.....A......A...C.....T...G.A.....T...T.A-AC.......C....T..A..C...
Batrachemys          C-............T.AT....T.........T..........A-..T..T...C.C........C..C
Sternotherus         C-.......C......A...T.....T.......A...T.....A-.........TC....T.........
Alligator            C-..C.GC..T......T....CC.A........G.....TA.A-....GC...C.......G..T...
Caiman               .-..C.G...T..C.....T..CC.A.....T.........A.A-..C..C...C....T..A..C..C
Tantilla             .-A...G.AAC..C..AACC..TA....T....AG..T.....A-..C..C..TC....TA.T.....C
```

Nucleotide conservation

One of the common features of the *nad3* frameshift site in turtles appears to be the use of two rare serine codons immediately following the inserted nucleotide, put in-frame as a result of the insertion. These two codons, AGT followed by AGC, are conserved in all of the turtles (as well as all sequenced birds) that carry the frameshift (Figure 10), but the corresponding nucleotides are frequently variable in those that do not (Figure 11). The AGT appears to be the required stall-inducing rarely used codon, and the conservation of the AGC suggests they both may have a role in frameshifting.

A second conserved feature found in all turtles and birds that carry the frameshift is a leucine codon as the last conserved 0-frame position, the codon that is at the P-site of the ribosome where the shift is thought to take place. This codon is CTB in all sequenced organisms with the frameshift, where B is the extra nucleotide and is either T, G, or C. It is the third position of this codon that disrupts the reading frame, and may be the inserted nucleotide. Regardless, this last position of the codon needs to be skipped in order to maintain the conserved amino acid sequence of the *nad3* polypeptide. In the sequenced turtles, this nucleotide is most often a C, occurring 9 times, but there are also three instances of T in this position and two Gs. It does not seem to be important which nucleotide is inserted, so long as it is not an adenosine. Analysis of the same region in all complete avian mitochondria in the database shows that the nucleotide in this position is always a cysteine, which is in keeping with the fact that birds have closer evolutionary relationships to each other than do the more divergent groups of turtles. The evidence against an A in the third position of the leucine codon extends beyond the 13 sequences from this study to all known examples of the *nad3* frameshift in birds and turtles, with the notable exception of the Reeve's turtle (*Chinemys reevesi*) (Nie, Pu and Peng, unpublished, NV-006082). I was not able to obtain samples of this species to verify the sequence in this region. Beyond this, only organisms that do not require a frameshift to translate *nad3* use an A in the third codon position. That the CUA codon is not usually found in organisms requiring the frameshift may be due to it being a perfect match for the tRNA-Leu (anticodon UAG) that recognizes the CUN codons. This strong binding may not allow for the required level of frameshifting in most organisms.

Figure 10 **All sequenced turtles with the frameshift insertion**

(A) All sequenced organisms that have the extra nucleotide present in their *nad3* gene. Codons are spaced in triplets entirely in the 0-frame. The leucine codon is the last conserved residue, and the two consecutive serines result from reading in the incorrect frame due to the insertion.

(B) Shows the same set of sequences "modified" to read as codons in the original frame without the extra nucleotide. Note that these all show the highly conserved amino acid sequence found in this region in chelonians without the frameshift. Also of note is the AGA codon used in *M. parkeri*. This codon is thought to be a stop in vertebrate mitochondria, and as we can see here, in all other turtles with the frameshift nucleotide, it is a highly conserved arginine. Since the codon is in-frame, it is unlikely a frameshift occurs, as it likely does in *P. subrufa* frameshift sites. It is more probable that this is a case of redefinition, changing a stop codon to be read as a sense codon, in this case, arginine.

	A. Decoded in the original 0-frame	B. Translation over the extra nucleotide
Apalone	CGC TTC TTC CTC AGT AGC AAT TCT R F F L S S N S	CGC TTC TTC CTC A GTA GCA ATT R F F L V A I
Chelydra	CGA TTC TTC CTC AGT AGC AAT CTT R F F L S S N L	CGA TTC TTC CTC A GTA GCA ATC R F F L V A I
Geochelene carbonaria	CGA TTC TTT CTC AGT AGC CAT CCT R F F L S S H P	CGA TTC TTT CTC A GTA GCC ATC R F F L V A I
G. sulcata	CGA TTC TTC CTC AGT AGC AAT CTT R F F L S S N L	CGA TTC TTC CTC A GTA GCA ATC R F F L V A I
Indotestudo	CGA TTC TTC CTC AGT AGC AAT CTT R F F L S S N L	CGA TTC TTC CTC A GTA GCA ATC R F F L V A I
Macroclemys	CGA TTT TTC CTC AGT AGC AAT CTT R F F L S S N L	CGA TTT TTC CTC A GTA GCA ATC R F F L V A I
Rhinoclemmys	CGA TTT TTC CTC AGT AGC AAT CTT R F F L S S N L	CGA TTT TTC CTC A GTA GCA ATC R F F L V A I
Terrapene	CGA TTC TTT CTC AGT AGC AAT CTT R F F L S S N L	CGA TTC TTT CTC A GTA GCA ATC R F F L V A I
Malaclemys	CGA TTC TTC CTG AGT AGC AAT CTT R F F L S S N L	CGA TTC TTC CTG A GTA GCA ATC R F F L V A I
Trachemys	CGG TTC TTC CTG AGT AGC AAT CTT R F F L S S N L	CGG TTC TTC CTG A GTA GCA ATC R F F L V A I
Clemmys	CGA TTC TTC CTT AGT AGC AAT CTT R F F L S S N L	CGA TTC TTC CTT A GTA GCA ATC R F F L V A I
Mauremys	CGA TTC TTT CTT AGT AGC AAT TTT R F F L S S N F	CGA TTC TTT CTT A GTA GCA ATT R F F L V A I
Macrochelodina	AGA TTC TTC CTT AGT AGC AAT CCT * F F L S S N P	AGA TTC TTC CTT A GTA GCA ATC * F F L V A I

While there are other areas of genome sequence around the frameshift region that are conserved, it is often difficult to distinguish between nucleotide sequence conservation that could be related to frameshifting and protein sequence conservation. In the strictest sense, to hypothesize that a particular sequence influences frameshifting, it would have to be completely conserved in all organisms that carry the frameshift insertion, and at the least variable in those that do not. These criteria are likely too

stringent, as it is probable that slight variations in a particular sequence could still have the required frameshift stimulating properties, and conversely, it is also possible that frameshift stimulating elements remain with no detriment in organisms where the frameshift insertion is not present. Ideally, these areas could be tested in vivo, as in similar studies in yeast and *E. coli*. Unfortunately, with only sequence data and no experimental data from sequence manipulation, it is necessary to use these criteria to postulate that a particular sequence is important in frameshifting.

Figure 11 **All sequenced reptiles without the frameshift insertion.**

All organisms sequenced that do not have the extra nucleotide present in their mitochondrial *nad3* gene. Here we can see that this region is far more variable than it is in organisms that require a shift over a frameshift insertion.

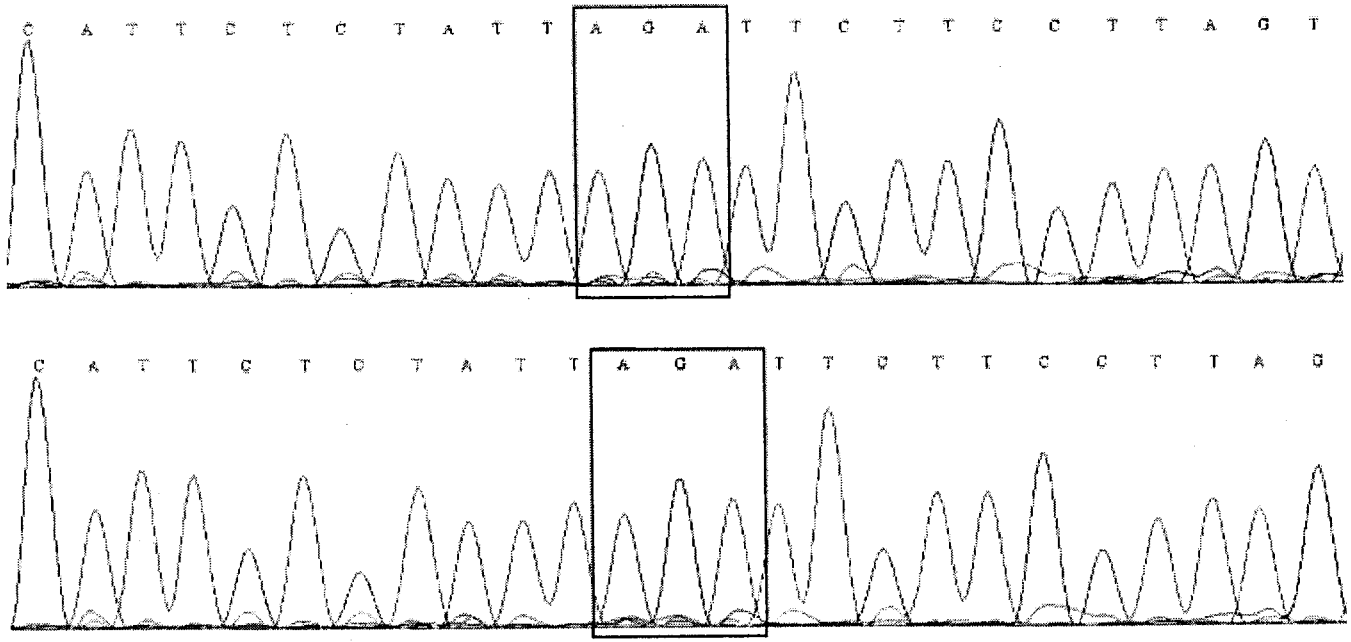
Pelomedusa	CGA	TTT	TTC	CTT	ATC	GCA	ATC	TTA
	R	F	F	L	I	A	I	L
Staurotypus	CGA	TTT	TTC	TTA	ACA	GCA	ATC	CTA
	R	F	F	L	T	A	I	L
Batrachemys	TGA	TTC	TTC	CTA	GTT	GCT	ATC	CTC
	W	F	F	L	V	A	I	L
Sternotherus	CAA	TTT	TTC	CTA	GTA	GCA	ATT	CTA
	Q	F	F	L	V	A	I	L
Alligator	CGG	TTC	TTT	ATA	GTA	GGC	ATC	CTA
	R	F	F	M	V	G	I	L
Caiman	CGA	TTC	TTC	ATA	GTC	GCC	ATC	CTA
	R	F	F	M	V	A	I	L
Tantilla	CAG	TTT	TTC	CTA	GTC	GCC	ATT	CTA
	Q	F	F	L	V	A	I	L

There is significant conservation at the protein level in the region surrounding the frameshift site. Only one amino acid is changed in all the sequenced turtles with the extra nucleotide in the region between nine positions upstream of the frameshift site to 21

positions downstream. Two changes at the nucleotide level, both caused by transversions, are found in the same species, Parker's sideneck turtle (*Macrochelodina parkeri*). The first site, at position 163 relative to *T. scripta*, is extremely interesting, replacing what is normally a CGN arginine with an AGA, which defined as a stop codon in the vertebrate mitochondrial code (Figure 12). At this site, 10 of the 12 other turtles with the frameshift use a CGA codon, with singular examples of CGC and CGG. In turtle species without the extra nucleotide, two also have CGA arginine codons, while in the two musk turtles, arginine is replaced by either tryptophan (TGA), as in the Mexican giant musk turtle, *B. nasuta*, or glutamine (CAA) in the common musk turtle, *S. odoratus*. It is worth noting that besides being the only two turtle species to show amino acid substitutions at this position, they also do not have the frameshift insertion. The four species sequenced that were shown to lack that insertion also have five other amino acid substitutions in this area, two in each of *P. subrufa* and *S. triporcatus*, and an additional one in *B. nasuta*. This region of the *nad3* gene is quite conserved regardless of the presence of frameshift insertions, though it appears that selection is relaxed somewhat in the absence of a need for frameshifting.

Figure 12 **Trace files for the AGA codon present in-frame in the *nad3* gene of *M. parkeri***

Shown are two trace files from two different PCR products. The AGA codon outlined is present and unambiguous in both. This was also confirmed by the sequencing of the opposite strand (not shown).



African sideneck *nad3* frameshift

In one species without the common *nad3* insertion at position 174, there is a different insertion mutation further upstream. The addition of either a C or a T between positions 133-135 in *P. subrufa*, first recorded by Zardoya and Meyer, 1998, results in an AGA stop as the next downstream codon. The sequencing of the *P. subrufa* complete genome reveals that the last in frame codon is CUU, which is another example of a wobble-matched CUN codon decoded by the tRNA-Leu(CUN) (anticodon UAG). I was able to confirm this sequence independently from a specimen from the Empire of the Turtle in Florida. Unlike in *M. parkeri*, where an AGA stop codon appears to be redefined as a sense codon, in *P. subrufa*, the AGA must induce a frameshift to allow for accurate decoding of *nad3*.

Secondary structure analysis

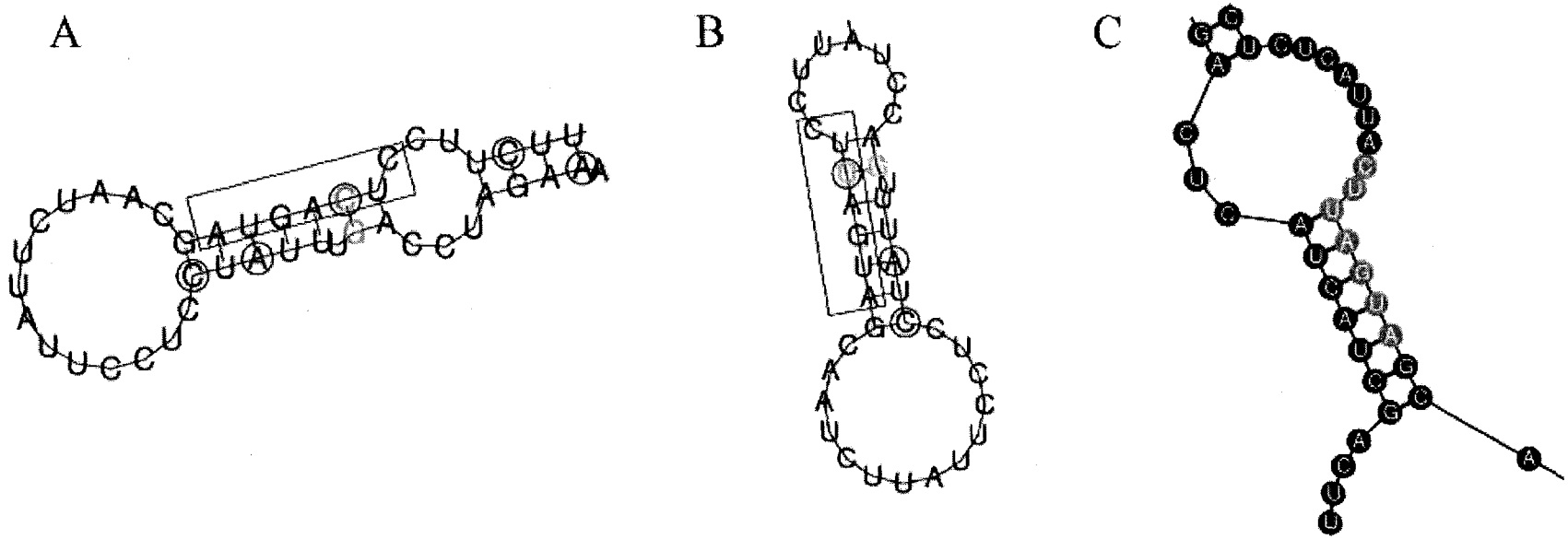
Using the Alifold web server, secondary structures were made comparing groups of sequences that contained the frameshift insertion at the conserved *nad3* site to those that do not. A variety of overall structures were obtained, and these are likely variable from one species to the next. The structures showed no overall consensus, but all groupings of chelonians with the frameshift at position 174 in the *nad3* gene display a stem-loop structure with a stem that is 7 base pairs in length and a 14 nucleotide loop. The stem structure involved the UCAGUAG sequence of the CUN AGU AGC A frameshift motif (Figure 13-A and 13-B). mFOLD analysis of the *nad4l* gene of *T. scripta* predicted a similar structure (Figure 13-C). The only known example of a +1 programmed translational frameshift that uses a known secondary structure element is the

mammalian antizyme. There is little similarity here however, as the pseudoknot in the antizyme gene is 3' of the frameshift site, while here the frameshift site is in the stem of the stem-loop, though the two structures may still have similar function.

The mFOLD results for *T. scripta nad3* also show a stem structure of six nucleotides, using the GAGUAG of the frameshift, though the loop here was highly variable. The results from *nad4l* placed the sequence UAGUAGC in a double helix region in all potential secondary structure conformations. The MARNA web server did not produce any conserved secondary structure for any of the groups of sequences used.

Figure 13 **Examples of possible stem-loop structures in genes with the frameshift nucleotide.**

(A) Consensus structure produced by the Alifold web server using the Vienna RNA Secondary Structure Prediction package for an alignment of small nucleotide fragments around the frameshift site from all sequenced chelonians that had the conserved *nad3* frameshift insertion at position 174 relative to the *T. scripta nad3* gene. The frameshift heptamer is enclosed in the rectangular outline. (B) Putative secondary structure produced by the Vienna package on an alignment of the complete *nad3* genes from all GenBank chelonian complete mitochondrial genomes that had the common *nad3* insertion along with the *T. scripta nad3* from this study. Again a rectangular outline is used to highlight the frameshifting heptamer. With the exception of the first nucleotide, the entire frameshifting sequence appears to be involved in the stem of a stem-loop structure. Similar analysis with sequences that did not have the extra nucleotide produced variable results, with no conservation of secondary structure around the frameshift site. (C) When analyzed with the mFOLD RNA secondary structure prediction algorithm, the complete *T. scripta nad4l* gene produces a number of different possible structures, but all conserve a stem region of complementarity between most of the frameshift site and the 5' end of the gene. The frameshift region is represented here in grey. Using the Vienna package on an alignment of the chelonian *nad4l* genes in GenBank predicted only the last three nucleotides to be involved in a stem structural element.



Complete mitochondrial genome sequence of the red-eared slider

Functioning as the cellular respiratory centre, mitochondria have undergone evolutionary pressure to minimize their genomes (Kurland, 1992a). They have retained their own translation system with mitochondrially-encoded ribosomal RNA and minimal complement of tRNAs, but all other ribosomal proteins are encoded elsewhere and are imported. Every one of these organelles has its own circular genome, inherited maternally. In animals, the genome is typically 16,000 to 20,000 base pairs in size, containing 13 protein coding genes, 2 ribosomal RNAs, and 22 transfer RNAs. It is these 22 tRNAs that are responsible for the translation of all 13 mitochondrially encoded proteins. The proteins are the adenine triphosphate (ATP) synthase (ATPase) subunits 6 and 8 (*atp6* and *atp8*), cytochrome b oxidase (*cytb*), cytochrome oxidase subunits 1-3 (*cox1-3*), and the NADH dehydrogenase subunits 1-6 and 4L (*nad1-6, nad4l*). The typical mitochondrial genome also includes a control region or non-coding region of variable length. The translation system of mitochondria has undergone a number of changes compared to from the standard rules of decoding, with programmed translational frameshifting being one of the most drastic. The complete *T. scripta* mitochondrial genome was sequenced to investigate this in chelonians.

The *T. scripta* mitochondrial genome contains the usual complement of mitochondrial genes in the standard arrangement for vertebrates. Upon analyzing the completed sequence, we also found a second frameshift mutation in the *nad4l* gene (trace files for each of these insertions can be seen in Figure 14). Here we will present a brief analysis of the features of the *T. scripta* mitochondrial genome, along with a more

detailed examination of both the common *nad3* frameshift as well as the newly discovered *nad4l* insertion.

Sequence annotation and analysis

The red-eared slider mitochondrial genome conforms to the typical vertebrate mitochondrial genome arrangement (Figure 15). It is comprised of 16,810 base pairs, and contains all 13 protein coding genes, 22 tRNAs, and 2 ribosomal RNAs normally found in vertebrate mitochondria. Nucleotide composition is 34.3% A, 25.9% C, 12.9% G, and 27.0%T, making it 61.3% AT and 38.7% CG.

Figure 14 **Trace files for the two *T. scripta* mitochondrial genome frameshift insertions**

In the three traces, the extra nucleotide is contained within a rectangular outline. The first trace is the *T. scripta* version of the common *nad3* frameshift insertion seen in many other turtles and birds. This site uses a G as the last nucleotide of the CUN leucine codon, observed in only one other organism. The second two traces are of the region surrounding the novel *nad4l* frameshift site. These are taken from two different PCR products, and in both, sequencing the opposite strand (not shown) confirmed the insertion. The only difference between these two frameshift sites is the last position of the leucine codon, a T in the *nad4l* site and a G in the *nad3* site.

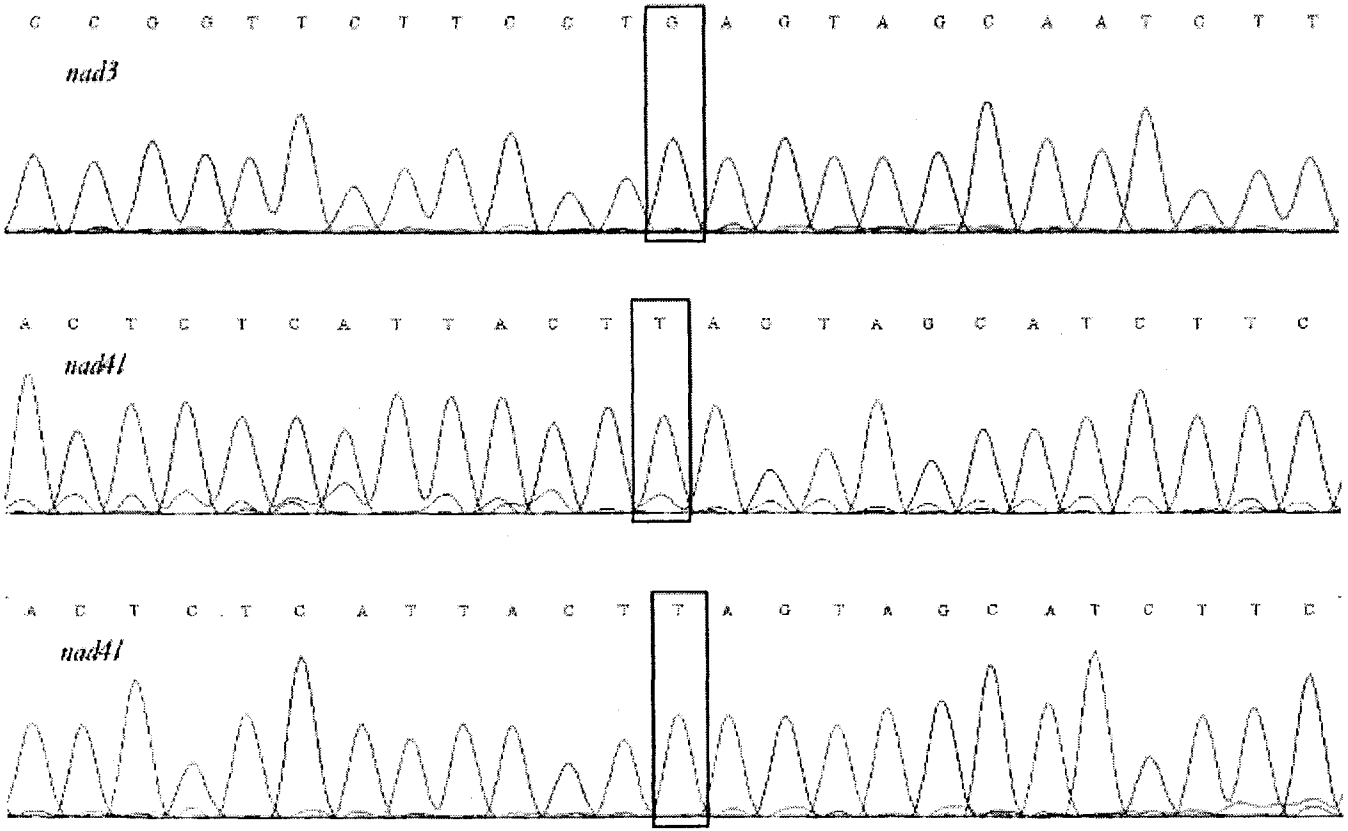
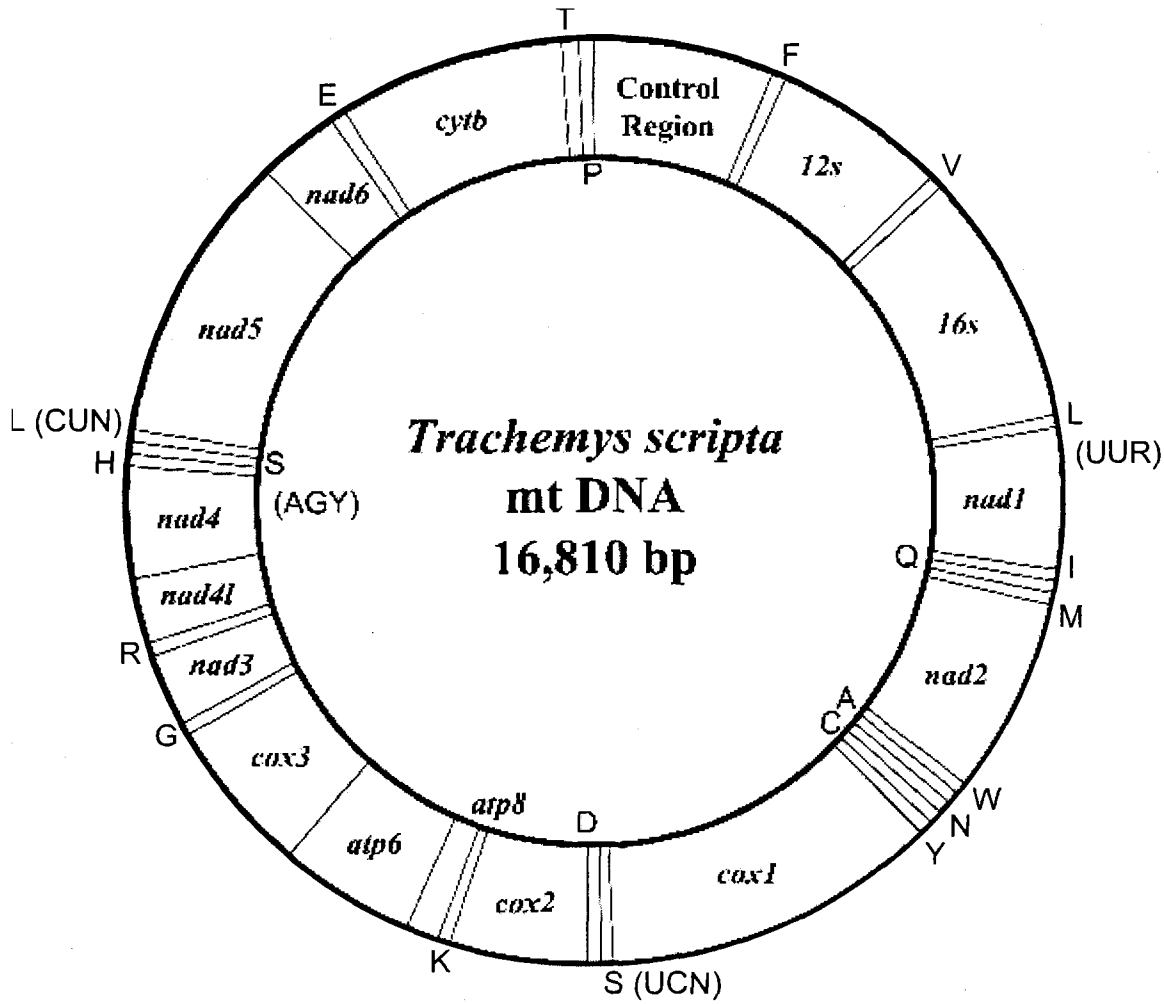


Figure 15 **Complete mitochondrial genome arrangement of *T. scripta*.**

A representative map of the complete mitochondrial genome of the red-eared slider turtle, *T. scripta*. All ribosomal RNA, tRNA, and protein coding genes are labelled. For blocks of tRNAs (for instance, the WANCY region immediately after *nad2*), labelling begins on the outside and then alternates inside to out.



Comparisons to other sequenced chelonians

The red-eared slider mitochondrial genome, compared to an alignment of *Chelonia mydas*, *Pelomedusa subrufa*, *Pelodiscus sinensis*, *Dogania subplana*, *Chinemys reevesi*, and *Chrysemys picta*, reveals some minor differences, but no major unique features. These differences are described in the relevant sections below.

Protein-coding genes

The 13 protein coding genes all align well with previously reported turtles. Eleven of the genes translate normally while two have frameshift insertions that disrupt the reading frame (Table 3). The *nad3* gene contains the inserted nucleotide at position 174 previously reported in other species, though the inserted nucleotide is likely a G, which is used far less than a pyrimidine at this site. A second frameshift insertion in the *T. scripta* mitochondrial genome is present in the *nad4l* gene, where what is likely a C or a T is inserted somewhere between nucleotide position 231 and 236 near the 3' end of the gene. Attention must be drawn here to the fact that the sequence immediately downstream of this novel frameshift site is identical to the sequence downstream of the conserved and original *nad3* frameshift site for seven nucleotides (AGTAGCA). It would appear that both the in-frame and +1 frame codons are important for frameshifting, as they are conserved at two different sites in *T. scripta*.

Table 3 **Ribosomal and protein coding genes within the *T. scripta* mitochondrial genome.**

A complete list of all rRNA and protein coding genes within the *T. scripta* mitochondrial genome. The positions of the first and last nucleotides are given, along with the total number of nucleotides. For protein coding genes, start and stop codons and number of amino acids are also given. For stop codons, T + AA or TA + A represents stop codon completion via the addition of adenosine nucleotides in during polyadenylation of the mRNA transcript.

Gene	Begins	Ends	Nucleotides	Amino Acids	Start Codon	Stop Codon	Notes
12S	71	1038	968	N/A	N/A	N/A	
16S	1111	2730	1620	N/A	N/A	N/A	
nad1	2806	3776	971	323	ATA	TA + A	
nad2	3986	5024	1039	346	ATA	T + AA	
cox1	5407	6951	1545	514	GTG	AGA	AGA stop is located in a five nucleotide overlap with tRNA-Ser (CUN)
cox2	7090	7776	687	228	ATA	TAA	
atp8	7852	8040	189	62	ATG	TAA	
atp6	8010	8693	684	227	ATG	TAA	Last A of TAA stop is shared between *atp6* and *cox3*
cox3	8693	9476	784	261	ATG	T + AA	
nad3	9546	9895	350	116	ATA	T + AA	Frameshift insertion at nucleotide position 174
nad4L	9966	10263	298	98	ATG	TAA	Contains frameshift insertion at position 234
nad4	10257	11637	1381	461	GTG	T + AA	
nad5	11847	13673	1827	608	ATG	TAA	
nad6	14325	13801	525	174	ATG	AGG	Only protein gene coded on the light strand
cytb	14403	15542	1140	379	ATG	TAA	

Transfer RNA genes

The *T. scripta* mitochondrial genome has the usual vertebrate mitochondrial complement of 22 tRNAs. This complement includes at least one that corresponds to every amino acid, with two tRNAs decoding the codons for serine and leucine. The tRNA-Ser (AGY) is of particular interest here, as it is responsible for the decoding of the consecutive AGY codons just after the frameshift site, should the ribosome not shift to the +1 frame. As in all other animal mitochondria, this tRNA does not fold into the standard cloverleaf structure observed generally for tRNAs (Figure 16). The codon immediately prior to the consecutive AGY codons at either frameshift site is a CUN leucine, also found at most other known mitochondrial frameshift sites, with CUG in *nad3*, or CUU in *nad4l*. These are decoded by tRNA-Leu (CUN) with an anticodon of UAG, which wobble pairs with the CUG codon in the third position, and is a mismatch for the same position at the *nad4l* CUU. In other organisms, this last in frame codon is rarely CUA, which would be exact Watson-Crick base pairing to the anticodon. It is possible that cognate codons in the peptidyl site for the leucine tRNA anticodon UAG are unable to promote required levels frameshifting. This poses a dilemma with regards to any proposed mechanism. It appears that the tRNA-Leu (CUN) that recognizes the last in frame codon is unable to shift in the +1 to read UGA, as a G-U wobble pairing in the first base and A-G mismatch in the middle base in the +1 frame are unlikely to be favoured over the original frame where the first two nucleotides pair via Watson-Crick with one mismatch in the wobble position. At the same time, codons that do not provide good binding to the leucine tRNA seem to be selected for at these frameshift sites.

Figure 16 **Proposed secondary structure for *T. scripta* mitochondrial tRNA-Ser (AGY) and tRNA-Ser(UCN).**

(A) The genetic sequences, foldings, and anticodons are listed for both tRNA-Ser (AGY) and tRNA-Ser(CUN). (B) Schematic representations of the divergent tRNA secondary structure of tRNA-Ser(AGY) and canonical tRNA structure of tRNA-Ser(CUN). The major differences are a missing DHU-loop in tRNA-Ser(AGY), along with an overly long acceptor stem. Either of these two characteristics of tRNA-Ser(AGY) may cause the ribosome to stall over AGY codons, and increase the frequency of frameshifting at these sites.

A

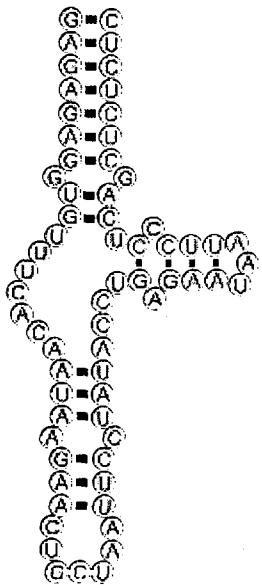
tRNA-Ser (AGY)

```
5'-GAGAGAGGUGUUUCACAAUAAGAACUGCUAAUUCCUAUACCUGAGAAUAAUUCCCUCAGCUCUCUC-3'
   1111111.11.......222.222..SER..222.222....3.333...333.3.11.1111111
```

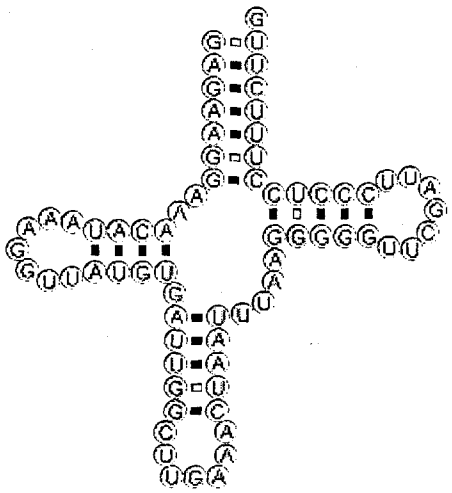
tRNA-Ser (UCN)

```
5'-GUUCUUUCCUCCCUUAGCUUGGGGGAAUUUAAUCAAAGUUCGGUUAGUGUAUUGGAAAUACAAAGGAAGAG-3'
   .111111122222.......22222....33333..SER..33333.4444.......4444..1111111
```

B



tRNA-Ser (AGY)



tRNA-Ser (UCN)

In both of the frameshift sites found in the *T. scripta* mitochondrial genome, two consecutive rare AGY codons require decoding via the unusual tRNA-Ser (AGY) (anticodon GCU), so both the AGU and the AGC (next two codons) have the potential to cause a pause. If the first of these two residues (valine) is not critical within the protein, it may give the ribosome two chances to frameshift. This may be another layer to the

mechanism that allows the extra nucleotide to be tolerated. While many frameshift sites employ an AGY codon as the first in frame codon after an insertion that is to be skipped over, all turtles and birds use the two AGY codons, always AGT followed by AGC. As noted, this sequence was found in both instances of frameshift mutations in the *T. scripta* mitochondrial genome.

Ribosomal RNA genes

The small ribosomal subunit (12S) is encoded from positions 71 to 1038, and is flanked by tRNA-Phe 5' and by tRNA-Val 3'. The tRNA-Val separates the 12S subunit from the large, or 16S, ribosomal subunit encoded between 1111 and 2730. 16S is punctuated at the 3' end with tRNA-Leu (TTR). Searches of the two mitochondrially-encoded rRNAs for sequences of high complementarity to the conserved frameshift motif revealed no large blocks of complementary sequence.

Control region

The control region is made up of a large repeat, a stretch of sequence of unknown function, and a TA microsatellite repeat. The large repeat is 89 nucleotides long, and is repeated once in its entirety with two substitutions at the 5 ' end, with a third partial repeat of the first 69 nucleotides. Following these repeats, there is an extended stretch of sequence with unknown function that is 798 nucleotides long. A small microsatellite AT repeat of 80 nucleotides separates the unknown region from tRNA-Phe.

Unassigned regions

The red-eared slider mitochondrial genome has a number of unclassified sequence elements (Table 4). Most are single nucleotides that are present between tRNAs and protein coding genes or other tRNAs. There is however a large block of sequence from positions 13,674 to 13,800 between *nad5* and *nad6* that has no apparent function, and returns no significant hits in a BLAST search (Altschul *et al.*, 1997), other than one to a previously sequenced *T. scripta nad5* gene (accession number 1216493). There is also a small region between tRNA-Asn and tRNA-Cys in the WANCY block of tRNAs that can potentially form a stem loop, which has led others to suggest that this might be the origin of replication for the light strand (Härlid, 1998).

Interestingly, the partial sequence present in GenBank of *T. scripta nad5* returned in the BLAST search of the unknown region between *nad5* and *nad6* appears to have another frameshift event (Cai and Storey, 1996) (Figure 17). This sequence includes a large coding block that normally shows high conservation, including two consecutive tryptophans, disrupted by an insertion, and subsequently restored downstream by a deletion. Though potentially interesting, especially as the insertion comes immediately prior to two AGN codons (AGT AGG), the existence of a similar frame-restoring deletion further downstream unfortunately casts major doubt on its validity. Also, as this is a well conserved region of the protein, it is unlikely that having such a long stretch of sequence out of frame would result in a functional protein. It is possible that the sequence was adjusted to restore the reading frame during the analysis, but sequencing error or the presence of a non-functional pseudogene is more likely the case.

Table 4 **All unassigned nucleotide sequence within the *T. scripta* mitochondrial genome.**

Nucleotide position of all unknown sequences. Most unknown sequence is single nucleotide spacers between tRNA genes, though there are three longer stretches of sequence that have no known function. The unknown sequence is generally conserved, suggesting that most has some function. The sequence between tRNA-Glu and *cytb* is not conserved however, and it would appear that the initiation codon of the *cytb* gene has moved at least six nucleotides 3' when compared to other chelonians.

Position	Sequence	Location	Conservation
5100	T	tRNA-Trp to tRNA-Ala	Conserved
5170	A	tRNA-Ala to tRNA-Asn	Conserved
5244-5268	CCTTTCCCGCTCTCTAAAAAGCGGG + AAA of tRNA-Cys	tRNA-Asn to tRNA-Cys	Conserved, exception is the African sideneck – may be origin of replication for light strand
5406	T	tRNA-Tyr to *cox1*	Conserved
7018-7019	CC	tRNA-Ser to tRNA-Asp	Partially conserved
7777	C	*cox2* to tRNA-Lys	Partially conserved
7850-7851	TT	tRNA-Lys to *atp8*	One T is conserved
11846	C	tRNA-Leu to *nad5*	Only *P. subrufa* has nucleotides here
13674-13800	127 nucleotides	*nad5* to *nad6*	Two other species have insertion but no sequence similarity
14394-14402	ACCAGACCA	tRNA-Glu to *cytb*	Not observed in other species
15543-15545	AAC	*cytb* to tRNA-Thr	Partially conserved
15620	G	tRNA-Thr to tRNA-Pro	Not conserved

Figure 17 **Alignment of a putative *nad5* frameshift site in *T. scripta* from a previous study.**

When aligned to the complete genome sequence, the *nad5* sequence from Cai and Storey's 1996 study appears to have a third example of the programmed translational frameshift site. The insertion appears immediately 5' of two consecutive AGN codons. There is a compensating deletion 44 nucleotides downstream also in red that restores the conserved reading frame.

nad5 from Cai and Storey 5'-AGGTAGTAGGAATCATATCCTTCTTACTCATTGGATGGTGACGCGGC-GAGAAG-3'

nad5 from this study 5'-AGG-AGTAGGAATCATATCCTTCTTACTCATTGGATGGTGACGCGGCCGAGAAG-3'

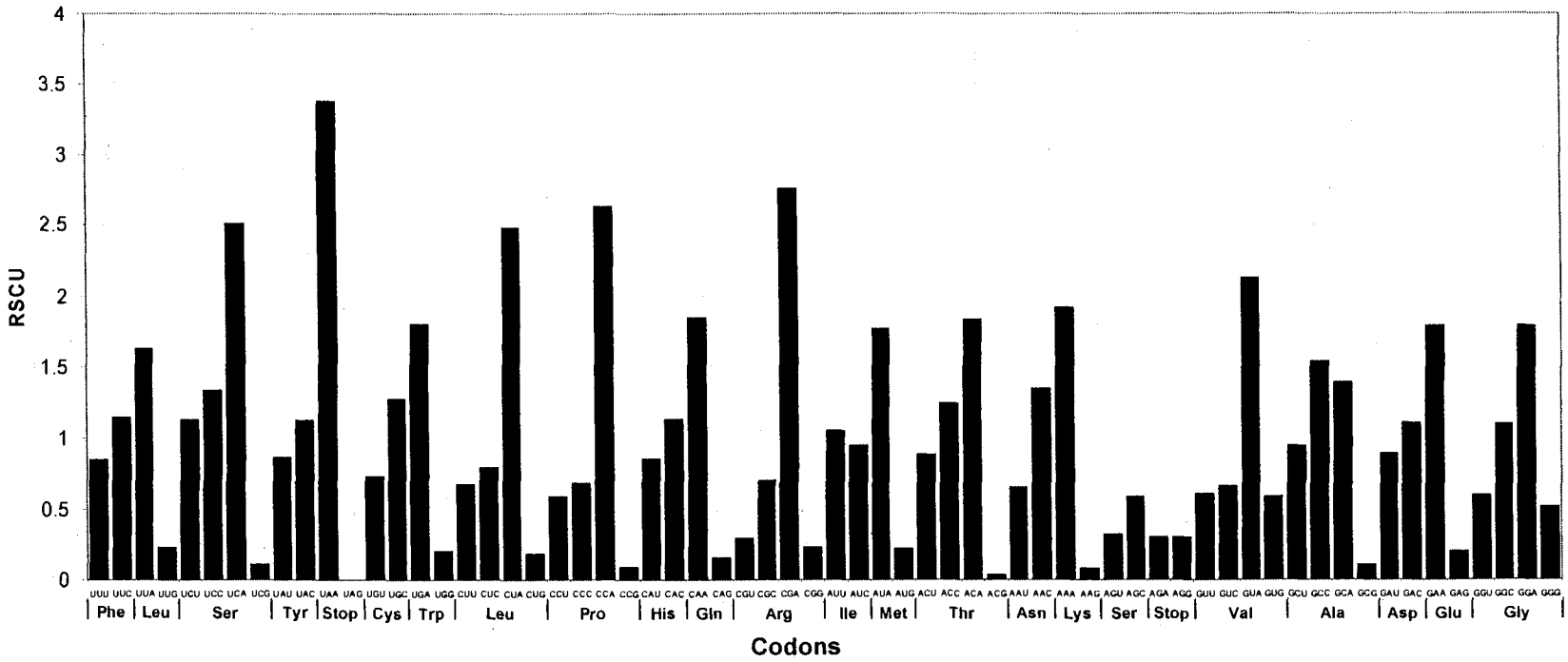
Codon usage analysis

Relative codon usage, relative synonymous codon usage, and overall codon usage were calculated for both the *T. scripta* mitochondrial genome and the chicken complete mitochondrial genome (Figure 18).

The frameshift heptamer CUB AGT A was not found anywhere else in-frame in the *T. scripta* mitochondrial genome. The AGTAGCA motif seen downstream of the frameshift insertion was also not found anywhere else in-frame, nor for that matter were any two consecutive AGY codons. Though there are 28 instances of consecutive serine codons, none had more than one AGY codon, and this was always in the second position. The only two places these sequences exist in the *T. scripta* genome are the two programmed frameshift sites.

Figure 18 **Relative synonymous codon usage in the *T. scripta* mitochondrial genome.**

All possible codons are represented, grouped according to amino acid or stop codon. Although the AGY serine codons are somewhat rarely used, if an AGY codon was all that was required to stimulate a frameshift, more selection against their use would be expected.



DISCUSSION

The two +1 frameshift sites in the *T. scripta* mitochondrial genome

Sequencing the complete mitochondrial genome of *T. scripta* revealed not only a conserved programmed translational frameshift site within the *nad3* gene, but also a novel frameshift site within *nad4l.* A similar situation appears in *P. subrufa*, where frameshift insertions were present at three different sites not found in other species. As in *P. subrufa*, there is high conservation between the different frameshift sites in *T. scripta.* In the *nad3* site, the conserved reading frame shifts +1 over the sequence CUG AGU A, written as codons of the original 0-frame. In *nad4l*, the change of frame occurs over CUU AGU A, and it would appear likely there are properties specific to this sequence that are essential in inducing the shift. Translation of either site gives the same result. In the 0-frame, they both translate as a leucine followed by two consecutive serines, while a leucine followed by valine and alanine is the protein sequence if the frameshift-causing nucleotide is skipped. The only difference between the two nucleotide sequences is the synonymous G or U in the wobble position of the leucine codon.

In the survey of the *nad3* frameshift, organisms with the insertion also showed complete conservation of the frameshift sequence found in *T. scripta*, with the solitary exception of a transversion of the final A to a C in *G. carbonaria*. This is strong evidence that in *T. scripta*, and likely in other turtles, the sequence of CUN AGU A stimulates +1 frameshifting. It also implicates the two relevant tRNAs, both tRNA-Leu (decoding

CUN) and tRNA-Ser (decoding AGY), as having roles in the frameshift mechanism of organisms where this sequence is present in-frame. Certain tRNAs have been shown elsewhere to have a major role in determining frameshift frequencies. For instance, in a study in the yeast Ty3 element, where GCG is used as the last in frame codon, mutating it to GCA – a change that causes it to be decoded by tRNA-Ala (UGC) rather than tRNA-Ala(CGC) – completely eliminates frameshifting (Vimaladithan and Farabaugh, 1994).

Comparisons to other +1 programmed translational frameshift sites

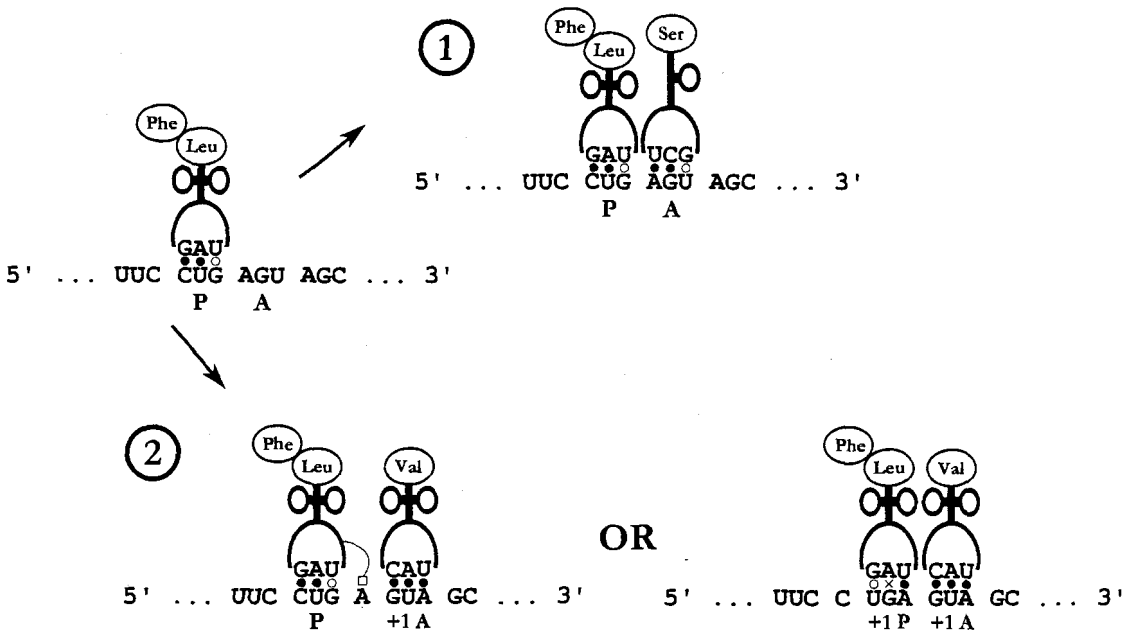
In the frameshift competent translation systems that are evidently present in organisms that require a shift in frames to accurately decode certain transcripts, the presence of frameshift-stimulating sequences and features likely promotes the necessary shift of reading frame. In the +1 programmed translational frameshifts in *E. coli prfB*, yeast Ty1 and Ty3 elements, and mammalian antizyme all have two such elements in common. The first is a pause in translation at the shift site, caused either by the slow decoding of a rare or nonsense codon in the next in-frame position and possibly aided by the presence of mRNA secondary structure. In yeast Ty3 elements, the frameshift heptamer is GCG AGU U. It is the AGU serine codon that is thought to cause the required stall allowing the ribosome to shift frames (Vimaladithan and Farabaugh, 1994). The AGU codon in the *nad3* and *nad4l* sites likely has a similar role. We believe this stall leads to a competition between a number of possible outcomes, from termination of translation to the frameshift required to produce a functional protein (Figure 19).

The second element is a peptidyl site codon that has poor wobble position pairing with the corresponding tRNA and often good pairing with the same tRNA if shifted +1.

Change to this position in known *E. coli* frameshifting genes alters frameshift efficiency by up to 1000-fold (Curran, 1993). Use of a common codon, or by extension one that is quickly decoded, in the +1 codon from the P-site codon has also been shown to aid frameshift efficiency (Hansen, 2003). Such is the case in the yeast Ty1 element and the majority of the frameshift sites found in the *Polyrhachis* ants. In that group of ants however, one site, TGG AGT A, does not have good +1 pairing for the P-site tRNA. This situation is similar in Ty3. In Ty3 elements the tRNA that decodes the first codon of the frameshift site GCG AGU U, the codon in the ribosomal P-site, is tRNA-Ala (GCN) (anticodon CGC), again with poor +1 binding (Vimaladithan and Farabaugh, 1994). In *T. scripta*, two different leucine codons are used in the equivalent position. In *nad3* it is CUG, and in *nad4L*, CUU, both of which are decoded by the tRNA-Leu (CUN). In both of the *T. scripta* frameshift heptamers, the P-site +1 pairing is poor.

Figure 19 **Possible outcomes after a ribosomal pause at the rare AGT codon in *T. scripta nad3*.**

In the decoding of the *T. scripta nad3* gene, we believe that, with tRNA-Leu (CUN) in the P-site, there are two possible outcomes as the ribosome attempts to read AGU. The first (1) shows canonical decoding of the AGU with tRNA-Ser (AGY). The unique structure of tRNA-Ser (AGY)is represented by the overly long schematic tRNA diagram. The structural differences between tRNA-Ser (AGY) and canonically-folding tRNAs may cause the proposed ribosome stall at this site. The second possibility is that the ribosome shifts frames to read in the +1. This may be possible in one of two ways, either by occlusion of the A-site (2) or by re-pairing of the P-site tRNA-Leu (3). As the *nad4l* frameshift site is nearly identical to that of *nad3*, a similar situation is likely for that frameshift event as well.



The role of the AGY codon and the tRNA-Ser (AGY) responsible for its decoding in +1 frameshifting

Use of AGY codons in animal mitochondrial +1 programmed translational frameshifts

Many examples of +1 programmed translational frameshifts have an AGY codon as the first codon of the 0-frame that is not decoded. This codon is present in frameshift sites from a diverse array of species, from the *Polyrhachis* ants to birds, and is again observed in this study in the two frameshifting sequence examples found in the *T. scripta*

mitochondrial genome as well as the 14 other chelonian *nad3* frameshift sites. With the exception of the three frameshift sites in the African sideneck turtle, *P. subrufa*, all avian and chelonian +1 programmed translational frameshifting sites have AGY in this position. In ants, the most common frameshift heptamer is GGN AGT A, with AGT likely causing the stall that shifts translation to the GTA valine codon. Further, the insertions are always immediately upstream of a GTA or a GCA, and four of five produce AGY as the first 0-frame codon after the insertion (Beckenbach *et al.*, 2005). In all examples, the AGY codon is thought to stall the ribosome allowing for translation to shift to the +1-frame.

Does AGU cause a stall in translation as a result of the slow recognition of the AGY codons by tRNA-Ser (AGY)?

In many examples of frameshifting, the ribosomal stall thought to be required is caused by a lack of availability of a particular tRNA. Indeed, studies have shown that limiting certain tRNAs has profound effects on frameshifting levels (Leipuviene and Bjork, 2005, O'Connor, 2002). In the case of rare codons contributing to frameshifting efficiency, rarity of the codon is often paralleled by a limited quantity of the tRNA that recognizes them. In human mitochondria, where the two groups of serine codons are decoded by two different tRNAs, tRNA-Ser (AGY) (anticodon GCU) and tRNA-Ser (UCN) (anticodon UGA), each tRNA is present in the mitochondria at relatively similar concentrations (King and Attardi, 1993). If this remains the case in chelonians – and conservation of mitochondrial function and constitution across species suggests that it would – a lack of the tRNA-Ser (AGY) is not the primary cause of the proposed frameshift-inducing ribosomal pause. It is likely that the stall is a result not from the lack

of availability of the tRNA-Ser (AGY), but from its comparatively slow rate of recognition of its cognate and near cognate codons (Hanada *et al.*, 2001), which would have the same effect as if the tRNA was rare. If the assumption that the levels of tRNA-Ser (AGY) and tRNA-Ser (UCN) is invalid, the alternative hypothesis that relies on the relative concentration differences can easily be made. In this case, lower levels of tRNA-Ser (AGY) in turtles may account for the chelonian predilection for frameshifting and its absence in humans and other mammals. Furthermore, tRNA-Ser (AGY) could be especially limited in chelonians, exacerbating the pause at AGY and allowing for higher levels of frameshifting, and in turn, increased tolerance for frameshift insertions prior to AGY codons.

Features of tRNA-Ser (AGY) related to +1 frameshifting

If the tRNA-Ser (AGY) is not limited in chelonian tRNA, there must be some features of this tRNA that cause AGY serine codons to be rarely used in comparison to UCN serines. These same features would likely allow it to produce the proposed ribosomal stall required for frameshifting. Most tRNAs have the same pattern of secondary structure and positions of universal nucleotides, a tenet that holds across kingdoms, from bacteria, archeabacteria, and eukaryotes to chloroplasts and plant mitochondria (Steinberg *et al.*, 1994). In animal mitochondria genomes, where tRNAs often differ from their chromosomal counterparts, tRNA-Ser (AGY) codons stands out as especially variable. This tRNA diverges greatly from this standard tRNA structure with no DHU-loop and an overly long acceptor stem (Arcari and Brownlee, 1980, Clary and Wolstenholme, 1985, Stewart and Beckenbach, 2003). The DHU-loop structural element is involved in tertiary interactions, and this therefore affects the structure of tRNA-Ser

(AGY). In order to remain functional, the tRNA-Ser (AGY) must be able to maintain a constant distance between the anticodon and the CCA terminus (Steinberg *et al.*, 1994, Watanabe *et al.*, 1994, Hayashi *et al.*, 1998). This is thought to be accomplished in one of two ways: either by folding into an L form (proposed by de Bruijn and Klug, 1983, Hayashi *et al.*, 1998), or into a boomerang shape (Steinberg *et al.*, 1994), with no current consensus. Length of the acceptor stem shortens over the evolution of animal species – reptiles and birds are closer in this regard to invertebrates – and both groups use AGY and AGN far more often than mammals. There appears to be a loose relationship here between acceptor stem lengths, how often the AGY serine codon is used, and how often frameshifting occurs. The shorter-stemmed tRNA-Ser (AGY) in mammals may have a reduced ability to decode AGY codons and as a result may be used less frequently. The various mitochondrial frameshift insertions are often present prior to these codons, and their low use in mammals compared to turtles and birds may explain the absence of any observed mammalian mitochondrial frameshift sites. It is apparent that the unusual structure of tRNA-Ser (AGY) has some role in programmed frameshifting sites that use AGY codons, though the mechanism by which it operates remains elusive.

Unlike tRNA-Ser (AGY), the second serine tRNA, decoding (UCN), forms a canonical secondary tRNA structure. In some ways, these two serine tRNAs are functionally equivalent. In an *in vitro* bovine mitochondrial system, both tRNAs were shown to have nearly equivalent abilities to form ternary complexes with mitochondrial EF-Tu and GTP. When tested for translational activity however, differences between the two tRNAs were revealed. tRNA-Ser (AGY) was shown to have a lower translational activity, producing only short polypeptides in a translational efficiency assay of at most a

tetramer in length (Hanada *et al.*, 2001). This is thought to be the result of either the tRNA-Ser (AGY) releasing from the mRNA from the P-site, or that the tRNAs are rate-limiting and do not enter and bind to the A-site at a sufficient rate. Operating again under the assumption that the relative concentrations of the two serine tRNAs in birds and reptiles are functionally equivalent as they are in humans, the tRNA-Ser (AGY) is a rate-limiting codon only as a result of these delays in decoding and not from its availability within the mitochondria. tRNA-Ser (AGY) was also shown to be disfavoured by the mitochondrial ribosome when there were other tRNAs present that could decode the same codon. With no other such tRNA within turtle mitochondria, the task is left to tRNA-Ser (AGY). The ribosomal stall required for frameshifting most likely arises from tRNA-Ser (AGY) having to overcome the difficulties in successfully decoding AGY outlined above.

The role of the last in-frame CUN codon in +1 frameshifting

Apparent selection against the CUA cognate codon at the ribosomal P-site

Experiments by Curran in 1993 varied the last in frame codon before the +1 frameshift in *prfB* in *E. coli* in an effort to determine the codons and tRNAs that allowed the required levels of frameshifting. This site is equivalent to the leucine CUN codon in the *T. scripta* frameshift sequence. In general, he found that both wobble pairing between the last in frame codon and the decoding tRNA in the third codon position along with stable pairing for the P-site tRNA and the first codon in the +1 frame greatly affect frameshifting efficiency. In the survey of the *nad3* frameshift site, none of the sequences that contained an extra nucleotide used cognate binding between the last in-frame codon position and the leucine tRNA. In the GenBank database, there is only one example, the

Reeve's turtle (*C. reevesi*) from all complete turtle and avian mitochondria that use CUA as the last in frame codon. Slippage, generally in the +1 direction, has been shown to occur over the CUB family of codons (Vimaladithan and Farabaugh, 1994). CUA, however, does not stimulate frameshifting – it is an exact cognate, making the case of the Reeve's turtle even more puzzling. We are then struck with a paradox – the P-site tRNA likely is unable to release its in-frame codon and re-pair in the +1-frame in the two *T. scripta* frameshift sites, yet it appears that there is selection operating in favour of non-cognate codons that are recognized by the tRNA-Leu at the P-site.

In order to conform to the currently proposed models for programmed +1 frameshifting, the tRNA in the P-site must do one of two things (Figure 19-3 and 19-3). The first possibility is that it breaks the codon:anticodon bonds in the zero frame and re-pairs in the +1 frame, a mechanism initially suggested by Sundararajan in 1999 and refined by Hansen in 2003. The range of pairs acceptable in the +1 frame has been shown to be greater than what is tolerated in the original selection of the amino-acyl tRNA in the zero frame (Curran, 1993), which would further enable this mechanism to stimulate frameshifting and may allow for the *T. scripta* sites to frameshift by this mechanism. This includes accepting G:U wobble pairing in 1st position of the shifted peptidyl tRNA in *prfB* (Craigen *et al.* 1985) and yeast Ty elements (Belcourt and Farabaugh, 1990). In the *T. scripta* frameshift in *nad3*, the tRNA-Leu (CUN) (anticodon UAG) would need to re-pair +1 with UGA to conform to this model. The resulting codon:anticodon complex requires a G:U weak wobble pair in the first position, a purine-purine mismatch (clash) in the second position, and a Watson-Crick pair in the third position (Figure 19-3). In the *nad4l* frameshift, the codon:anticodon complex has the same G:U wobble pair in the first

position followed by two A:T Watson-Crick base pairs in the second and third positions. It appears that in *nad4l*, the peptidyl-tRNA may be able to re-pair in the +1 frame, but a similar scenario at the *nad3* site seems unlikely.

The second model proposes that the P-site tRNA influences the incoming A-site tRNA to pair in the +1 frame. Three possibilities have been proposed to accomplish this (Vimaladithan and Farabaugh, 1994). The first proposes that the tRNA at the P-site is able to block a portion of the A-site anticodon. The second suggests that specific peptidyl-tRNAs are able to stabilize the interaction in the +1 frame, possibly by stabilizing the non-cognate tRNA binding, allowing it to remain after the EF-Tu kinetic proofreading step is complete. Finally, if normal peptidyl-tRNAs dissociate in the translational pause, there may be certain tRNAs that have the ability to remain hydrogen bonded and allow an eventual +1 continuation.

The effect of non-cognate codons on the ribosome during translation

We have shown that a cognate codon is seldom if ever used as the last in-frame at a frameshift site in animal mitochondria, a trend that extends to other examples of +1 programmed frameshifting. One explanation for this could be the different interactions that occur between cognate versus near cognate codons and tRNA anticodons, and the resulting effects on the ribosome. In bacterial translation, cognate tRNA codon recognition closes the 30S subunit (Rodnina *et al.*, 2005), resulting in the anticodon stem being tightly held by the closed 30S subunit. In the next step, following the more favourable course of action, the acceptor arm moves towards the peptidyl-transferase centre (Frank *et al.*, 2005). Near cognate tRNAs, however, cause the ribosome to behave

differently, which may be relevant to their ability to facilitate +1 frameshifting. The near-cognate pairing between codon and anticodon, such as between the CUB codon and the tRNA-Leu (CUN), allows the tRNA to leave the ribosome when the acceptor arm is free as it is only loosely bound at the anticodon stem-loop (Frank *et al.*, 2005). The destabilization of the aa-tRNA/ribosome complex caused by single mismatches between codon and tRNA anticodon are independent of where this mismatch occurs or what type it is (Rodnina et al, 2005). This helps to explain why in many frameshift studies, exact WC pairing at the last in-frame codon shows low levels of frameshifting generally. With cognate pairing, the tightly bound anticodon stem loop is unable to shift frames to the +1 position. Whether the anticodon stem loop is tightly bound may also play a role in programmed frameshifting events where the P-site tRNA does not change frames. In this case, either the loosely bound tRNA anticodon stem loop or a change in the conformation of the ribosome could affect frameshifting levels. With the prevalence of near-cognate CUB (B = C, G, or T) codons as the last in frame codon in the *T. scripta* and many other mitochondrial +1 frameshift sites, destabilization likely allows the leucine tRNA more readily to enable frameshifting.

Possibilities of a redundant frameshift mechanism

Most +1 programmed translational frameshift sites consist of a heptamer of nucleotides, made up of the P-site codon, skipped frameshift nucleotide and common codon in the +1 A-site position. In *T. scripta* and the other chelonians sequenced, it would appear that the next three nucleotides 3' are also important. In the original frame, these nucleotides result in a second AGY serine codon, AGC in the 0-frame followed by

an A (CUN AGT A → CUN AGT AGC A). This raises the possibility of a second frameshift site, shifting over AGU AGC A to be read as AGU A GCA if the shift from CUN AGU A to CUN A GUA does not occur. We investigated the possibility of the ribosome shifting frames at more than one site along the *nad3* mRNA transcript. If the CUN leucine and initial AGU serine are decoded accurately, a second ribosomal stall and frameshift may be possible at the second rare codon, which would again not be efficiently translated.

A precedent for a gene having more than one frameshift opportunity is found in an antizyme gene, antizyme 3 (Ivanov, 2000). The frameshift in this gene has no typical cis-acting frameshift stimulating elements, but it does have two 5' "pseudo-frameshift" sites that are nearly identical to the original frameshift site at the end of ORF1. It is possible that the ribosome is able to shift frames at any of these sites, and the result will be a functional protein. If this were the case, then the extra frameshift sites essentially give the translational mechanism two more chances to decode the message correctly. *E. coli* have been shown to undergo high levels of frameshifting over tandem AGG or AGA, or any combination (Spanjaard *et al.*, 1990, Spanjaard and van Duin, 1988), which may be the result of a similar mechanism where more than one possible frameshift can take place over the AGG AGG or AGA AGA sequence.

Though there is good precedent for genes encoding more than one frameshift site in an apparent effort to "hedge their bets", with indirect evidence for this occurring over consecutive AGN serine codons, it remains unlikely in *T. scripta* and other similar vertebrate mitochondrial frameshifting sites. Amino acid conservation in these regions is high, and if the first AGY codon was translated accurately, it would replace a highly

conserved valine present in nearly all vertebrates with serine. There is no way to unambiguously determine the frameshift site from the nucleotide sequence alone. To do so would require sequencing the protein, a task that remains for a future endeavour.

Comparing the frameshift sites of *T. scripta* and *P. subrufa*

Prior to our discovery of a second frameshift site in *T. scripta*, the only previously documented example of more than one frameshift site within a single vertebrate mitochondrial genome was in *P. subrufa* (Zardoya and Meyer, 2001). While the two frameshift sites found in the red-eared slider employ nearly identical sequences, these two sites differ greatly from the sites found previously in *P. subrufa*. Within each species however, the different frameshift sites are very similar. In *T. scripta*, the features are identical to the common *nad3* frameshift. The two sites conform to the CUN AGU A motif. In *P. subrufa*, the three sites are all variations on CUN AGA N, with a single shift required to translate *nad3* accurately and two for *nad4l*.

One copy of the putative *Euplotes crassus* protein 26 (pEC26) also needs at least two frameshifts, possibly three, for complete translation (Klobutcher, 2005), so the multiple frameshift sites in the *nad4l* gene of *P. subrufa* are not unprecedented. *P. subrufa* also has a (TA)-repeat microsatellite, as does *T. scripta*. Other shared features include a long non-coding region between *nad5* and *nad6*, an overlap between *atp6* and *cox3*, and the presence of extra nucleotides in *nad3* and *nad4l*. None of these similarities seem to explain why these two species are able to tolerate more sites prone to frameshifting than others, nor do their differences provide any insight into why the frameshift sequence is specific to each species.

Other +1 frameshift stimulators at the chelonian frameshifting sites

Other possibilities for frame maintenance and frameshifting must also be considered. A rarely used codon by itself is not enough in itself to cause increased frameshifting (Spanjaard *et al.*, 1990, Gallant and Lindsley, 1993). Combined with the presence of a slippery CUB leucine codon however, the downstream sequence of AGU A may be sufficient to cause the required level of frameshifting. Other possible frameshift stimulators include tRNA modifications and mRNA secondary structure. While we did not examine tRNA modifications, our secondary structure prediction analysis of the conserved *nad3* frameshift site revealed conservation of a stem loop structure in various groups of organisms with the frameshift insertion that involved the frameshift heptamer. Similar structure was not found in organisms without the extra nucleotide. The *nad4l* also had most of frameshift heptamer involved in a structural element. This may have some function in frameshifting, though how this would be accomplished is unclear.

Sequence, not structure, of a downstream element 14 nucleotides in length, stimulates frameshifting in Ty3 (Li *et al.*, 2001). This sequence is thought to interact directly with the ribosome. To date there remains no evidence for rRNA in vertebrate ribosomes interacting with mRNA, though rRNA-mRNA may interact in some initiation events (Raman, 2006) and searches of the complete chelonian mitochondrial *nad3* genes with the insertion do not reveal any apparent conserved sequences not present in *nad3* sequences without the extra nucleotide. The Ty3 stimulating sequence does not appear to have a functional analogue in the *nad3* frameshift. There is also a 50 nucleotide sequence 5' of the frameshift site in antizyme, stimulating frameshifting 2.5-5 fold (Matsufuji, *et al.*, 1995). This could function in a fashion similar to the Shine-Dalgarno sequence in

prfB. Again, there appears to be no equivalent sequence in any of the chelonian mitochondrial DNA sequences. Finally, in the antizyme frameshift, the 3' RNA pseudoknot also stimulates frameshifting, at a threshold of 2.5- to 5-fold. There does not appear to be any other features of either the *nad3* or *nad4l* genes that would suggest a role in frameshifting efficiency, but nothing can be determined with certainty without experimental analysis of the frameshift in a mitochondrial translational system similar to that for bovine mitochondria. Unfortunately, the analogous system is presently unavailable for chelonians.

Finally, we examined both the 12S and 16S ribosomal RNAs present in the *T. scripta* mitochondrial genome for any complementarity to regions near the frameshift site, which may be functionally equivalent to the Shine-Dalgarno sequences. No sequences displaying this property were found, though this is hardly unexpected. Any ribosomal influence on frameshift efficiency will likely involve not only primary sequence, but also secondary and tertiary structure, to say nothing of possible interactions with the ribosomal protein complement.

Possible origins for the avian and chelonian programmed mitochondrial +1 frameshifts

We propose that the extra nucleotide is an ancestral condition that first arose in a common ancestor of turtles and birds, and has been subsequently lost in various lineages. Using the newly discovered *nad4l* programmed frameshift *T. scripta* as an example, a possible mechanism for the origins of a programmed translational frameshift requires three intermediate states arising sequentially leading to the current state. The first

condition is the presence of the CUA GUA GCA motif, so that a CUN leucine codon, the frameshift stimulating codon present in all other chelonian and avian programmed mitochondrial frameshifts, is the last in-frame codon before the frameshift. Though the evidence from this study and experiments previously conducted provide much evidence implicating at least the CUB AGU A sequence as a stimulator of frameshifting, it may not be the only element required to allow a frameshift insertion to remain in a mitochondrial protein coding gene, and other sequences or structures may be required prior to insertion. A second condition involves the ribosome becoming more amenable to +1 frameshifting at certain sites. This seems likely considering that certain groups of organisms show a high propensity for mitochondrial frameshifting, while others show no evidence of the phenomenon. The last step is the insertion mutation of an extra nucleotide immediately upstream of A GUA GCA, resulting in the sequence CUN AGU AGC A, written in the original 0-frame with the **N** denoting the inserted nucleotide. With these elements present in the mitochondrial genome and translational system, we believe there is a sufficiently small negative effect on the organism to allow the frameshift to be tolerated and persist.

A similar argument can be made, in reverse, for the loss of the frameshift nucleotide observed in other chelonian species. We will use the example of the common musk turtle *S. odoratus* to illustrate how this loss could occur. The sequence around position 174, the common location of the frameshift nucleotide, in *S. odoratus* is CUA GUA GCA. Note that this is an organism that retains the A GUA GCA sequence motif, though, as shown by the spacing between the nucleotides, not in-frame. We suggest that this is an intermediate state between the presence of the extra nucleotide and modification

of the AGUAGCA motif through accumulation of synonymous mutations, and that it is potentially the result of two evolutionary events in series. The first event is a loss in efficiency in the ribosome's ability to frameshift at position 174 in the *nad3* gene of our example species, *S. odoratus*. If the efficiency loss is high enough to have negative effect on the fitness of the turtle, either a reversion to increased frameshift ability in the ribosome and retention of the extra nucleotide or a deletion mutation of the extra nucleotide would be favoured evolutionarily. Assuming a deletion occurred recently enough in evolutionary terms, the AGUAGCA motif would remain and would not have accumulated any mutations that, with the constraints put on it to be a frameshift promoter relaxed, are now allowed to accumulate. Thus, it would appear that the musk turtle demonstrates the intermediate condition, while species such as the Giant Musk turtle and the Toad-headed turtle, with changes to the AGUAGCA sequence, are examples of sequences that have either lost the extra nucleotide earlier, or alternatively have more rapidly accumulated substitutions after its loss.

AGA codon redefinition in the *M. parkeri nad3* gene

Are AGA and AGG redefined or even undefined codons in some vertebrate mitochondrial genomes?

In Parker's sideneck turtle, *M. parkeri*, there is an AGA codon in-frame in the *nad3* gene. Interestingly, this position is a highly conserved arginine residue in other species of turtle. AGA is normally classified as a stop codon in mitochondrial genomes, but is arginine when decoded by the standard code. These AGR codons appear only twice in-frame in the *T. scripta* mitochondrial genome, in both cases as stop codons. AGA terminates *cox1*, while the last codon of *nad6* is AGG. The presence of an in-frame AGA

in what is normally coding sequence of *nad3* raises the question of whether these are truly terminator codons in all vertebrate mitochondria, or if they can be redefined as sense codons in certain organisms. It is also possible that these codons are undefined in vertebrate mitochondria, and their decoding is a result of the different rates for a number of possibilities, rates which vary in between species.

Is the AGA codon an unassigned codon in some species?

In mammalian antizyme, the frameshift is stimulated by a UGA stop, the rarest of all the stop codons, and as such, may be the least efficient at recognition of translation termination. The resulting pause has been shown to stimulate frameshifting by 15 to 20 fold (Ivanov, 2000). In vertebrate mitochondria, the AGA and AGG codons are rarely used as stop codons. Though the rate of codon recognition is similar for cognate and near-cognate ternary complexes (Gromadski and Rodnina, 2004, Rodnina *et al.*, 1996), the AGG codon has been shown to induce a stronger pause than AGU in yeast (Vimaladithan and Farabaugh, 1994). If the AGR codons are similar in this regard, AGA may also produce an overly long ribosomal stall on the mRNA transcript. Two scenarios are possible here. The first is a translational frameshift event. This is likely the most favoured event in the *P. subrufa* mitochondrial genome when an in-frame AGA is encountered as two genes require frameshifts to be decoded accurately after encountering an AGA (Figure 20 B). The extended pause may also induce the ribosome to accept a non-cognate tRNA, which is what we propose occurs in the *M. parkeri* mitochondrial genome. The long pause may in effect force the ribosome to accept either the conserved tRNA-Arg (CGN) (anticodon UCG) or potentially the tRNA-Ser (AGY) (anticodon GCU) (Figure 20 A). Each tRNA has one G:A mismatch, tRNA-Arg (CGN) in the first

position, tRNA-Ser (AGY) in the third position, with two consecutive Watson-Crick pairs. An alternative to these suggestions is the use of an imported nuclear tRNA to decode the AGA codon in the *M. parkeri* mitochondria. This has been shown to occur in vertebrates (Dörner *et al.*, 2001), and could be an alternative possibility for the decoding of the AGA codon in *M. parkeri*.

Figure 20 **Possible outcomes for the decoding of the AGA codon in *M. parkeri* and *P. subrufa*.**

(A) Three possibilities are suggested for decoding the AGA codon in *M. parkeri*. In the first, a release factor recognizes the AGA codon and translation is terminated. In the second, tRNA-Arg (CGN) decodes the AGA codon, with a mismatch in the first position and two Watson-Crick pairs in the second and third. The third possibility is the decoding of AGA with tRNA-Ser (AGY), requiring the same mismatch as with tRNA-Arg (CGN), only in the third position. (B) Along with the three possibilities in *M. parkeri*, a fourth exists in *P. subrufa*. In order to produce the conserved amino acid sequence of *nad3* and *nad4l*, the tRNA-Leu (CUN) must affect a frameshift to the +1 frame to facilitate correct translation of the *nad3* and *nad4l* proteins. The *nad3* frameshift site is depicted below, and the frameshift may be caused either through re-pairing of the P-site tRNA-Leu, or occlusion of the first position of the A-site and out of frame binding by the incoming A-site tRNA-Glu. As this produces a conserved protein product, we believe it is favoured in *P. subrufa*.

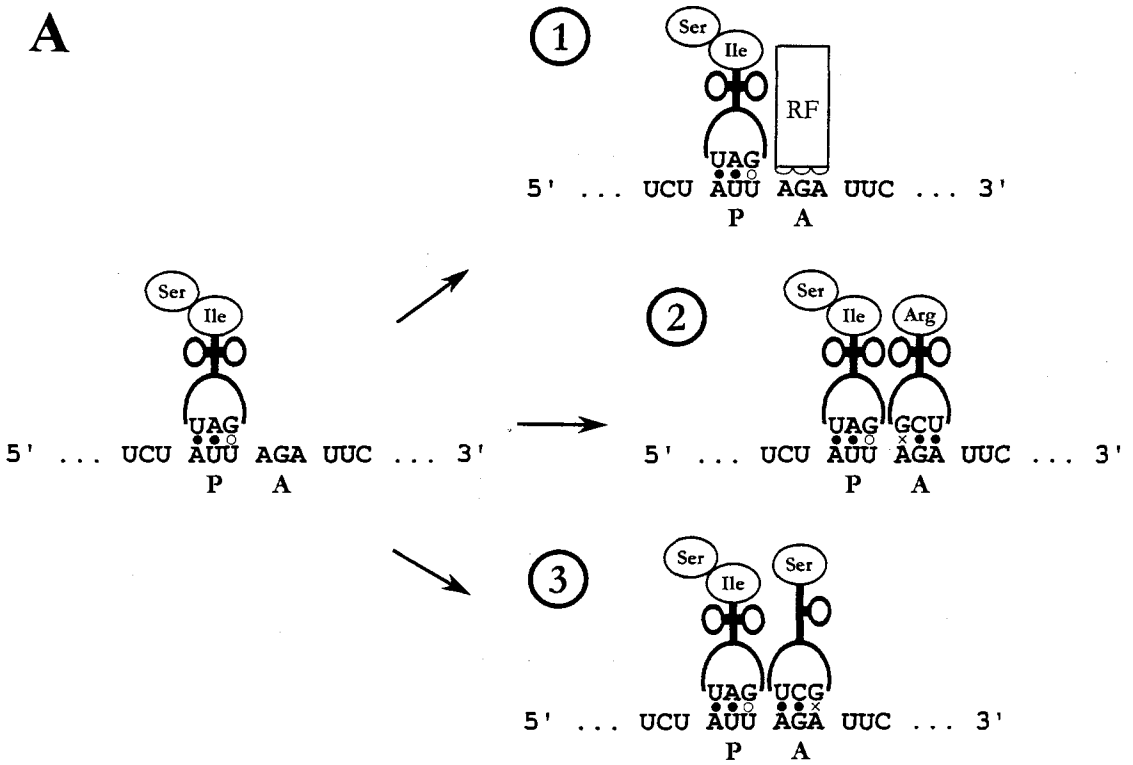
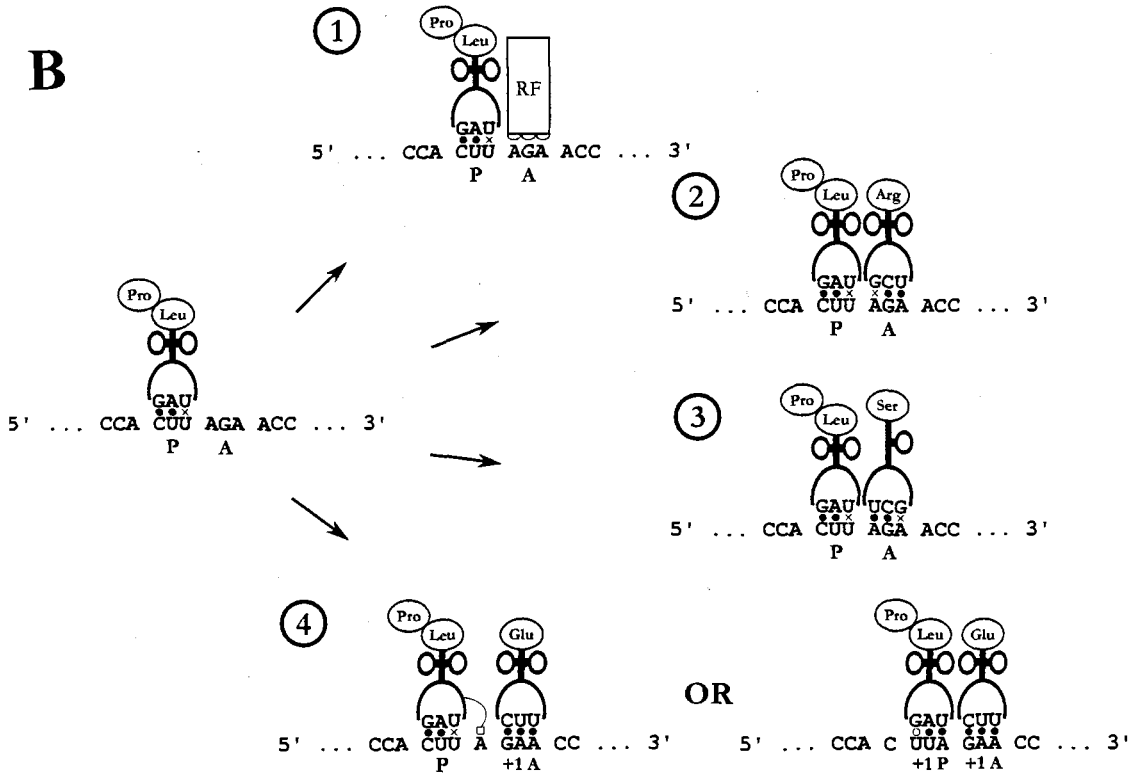


Figure 20 (continued)



Euplotes have a translation frameshift site that is either AAA TAA, or AAA TAG. This taxon has reassigned the UGA codon from a stop to encode cysteine (Kervestin *et al.*, 2001, Chavatte *et al.*, 2003). It has been suggested that as the *Euplotes* release factor 1 (eRF1) does not recognize UGA, it may also be slow to identify UAA and UAG, resulting in a stall in translation. This stop codon reassignment (or delay in recognizing release factors) may result in a pause in translation and then the ribosome shifting frames (described more completely in "Shifty Ciliates", Klobutcher and Farabaugh, 2002). In vertebrate mitochondria, the opposite scenario has occurred – the AGR codons that formerly coded sense codons have been redefined as stops. However, a similar ribosomal pause may occur when these codons are encountered, a delay resulting

also from codon redefinition and subsequent slow recognition. Interestingly, a study looking for tRNA-like structures within the ribosome that recognize terminator codons found evidence for two tRNA-like structures within the large subunit or rRNA, with anticodons complementary to AGA and AGG (Ivanov, 2001). The authors propose that these structures, which they call term-tRNAs, are responsible for the recognition of the AGA and AGG codons in vertebrate mitochondria. If these structures are responsible for terminating translation at AGG and AGA codons, perturbations to their structure in *P. subrufa* and *M. parkeri* may be the reason for the ribosome's apparently non-canonical behaviour at these sites. Sequencing of the *M. parkeri* ribosomal genes and comparison to *P. subrufa* and other chelonians rRNA genes may reveal some interesting results.

General frameshift site characteristics

Are frameshifts only tolerated in enzymatic/low production output proteins?

In *Euplotes*, the various frameshift sites appear only within genes that encode enzymatic products, or proteins with enzymatic function. Frameshifts are conspicuously absent from any abundant proteins (Klobutcher, 2005). If we expand this observation to include all known examples of programmed translational frameshifting, the principle remains – programmed frameshifting is seldom found in highly expressed genes. Of the known *Euplotes* genes that require a frameshift, six are enzymatic or enzymatically-associated proteins and the remaining two are of unknown function, although are not believed to be abundant. All of the known abundant *Euplotes* genes (27 in all), including tubulins, histones, and ribosomal proteins have no frameshifts (Klobutcher, 2005). From these numbers, if frameshifts were randomly present within the genes, it would have been

probable that some of these genes would show frameshifts, suggesting that frameshifts are perhaps less tolerable within highly expressed genes. Applying this argument to avian and chelonian mitochondria, we note that frameshifts appear to be relatively abundant in the NADH group of genes and in *cytb* genes, but have never been found in any of the *cox* set of genes. Again, we are restricted to a very small sample size as a result of the nature of the mitochondrial genome, but if frameshifts could be tolerated anywhere in the genome so long as the necessary contextual sequence was present, we would expect to find examples in a more diverse array of locations within the genome.

Tolerance of innocuous +1 frameshifts in mitochondrial genes

It has been suggested that moderate frameshifting during the expression of some genes is not bad enough to facilitate strong negative evolutionary pressure -- that some level of unintended frameshifting is tolerated (Gurvich *et al.*, 2003). A similar hypothesis is suggested by Klobutcher (2005) in his work on *Euplotes*. They both propose that some level of erroneous/incomplete/non-functional protein product is not detrimental enough to cause negative selection against the frameshifting heptameric sequences that cause them, at least in moderately expressed genes. In *E. coli*, none of the highest transcribed genes have frameshifting slippery sequences (Shah *et al.*, 2002). They postulate also that long genes would not tolerate any frameshifting sequences. While mitochondrial genes are short by virtue of their location, it is in some of the shortest genes that the majority of frameshifts are observed.

Lack of regulatory function for mitochondrial +1 frameshifts

Unlike many classical programmed translational frameshift examples, there is no evidence for regulation of genes through frameshifting in mitochondrial sites, nor is there for frameshifting genes in *Euplotes.* It appears that animal mitochondria and the *Euplotes* (possibly to a greater extent) tolerate inserted nucleotides at certain sites without having any known regulatory function – so long as they are selectively neutral, they are retained. Klobutcher cautions that a difficulty arises here in *Euplotes* in distinguishing between +1 frameshifts that have arisen for regulatory means and the organisms evolving frameshift tolerant translational machinery that efficiently translates over certain frameshifting sites. In mitochondria, this does not appear to be troublesome – if these frameshifts served a regulatory role, it would be extremely unlikely that they would then be secondarily lost in such a large proportion of organisms with no apparent detriment, though this does not entirely eliminate the possibility.

CONCLUSIONS

The translation systems of some vertebrate mitochondria allow them to frameshift at high frequency over the sequence CUB AGU A

Frameshift insertion mutations into genes happen with relative frequency. Normally the ensuing protein product would be non-viable and the insertion eliminated through natural selection. The only time they are observed is when they are tolerated and the gene can produce a full length, functional product. Where present in mitochondria, frameshifting over these sites does not appear to have a regulatory role, but is nevertheless tolerated under certain conditions. This is a result of a nucleotide insertion into a specific sequence that is required to promote frameshifting. In chelonian mitochondria, and animal mitochondrial genomes in general, this consists of an in-frame codon, almost always a CUB leucine, followed by a rare or non-sense codon that stalls the ribosome overlapped in the final two nucleotides by a more commonly used sense codon. The other stipulation is that this also requires that the mitochondrial translational machinery be amenable to frameshifting such that this context sequence produces enough functional protein products so there is minimal selective pressure against the frameshift insertion. In some turtles it appears that insertion mutations into CUA GUA GCA at the third position of the leucine codon appear to suffer little negative selection. No evidence for gene regulation through frameshifting was found in turtles, an observation that holds for all the known frameshift sites within animal mitochondria – they are tolerated, but do not appear to have regulatory function.

AGR is a redefined or unassigned codon in some vertebrate mitochondrial translation systems

An AGA codon is present in-frame in Parker's Sideneck Turtle at what is a conserved arginine residue in nearly all other sequences turtles. In the vertebrate mitochondrial code, AGR codons are considered terminators and have not been previously shown to be present in-frame in mitochondrial genes anywhere except as the last (terminator) codon, and even these are used infrequently. We propose that one of at least two events can happen at an in-frame AGA codon, with the possibility that these may extend to AGG as well:

1. If there is a tRNA which efficiently decodes AGA, translation continues until a terminating codon is reached that binds the release factor. This case occurs in nuclear genes using the standard code, where AGR codes for Arg. This would also occur if the codon was decoded in the mitochondria by an imported nuclear tRNA.

2. If there is no tRNA that decodes AGA, then a stall in translation occurs with the codon that is immediately upstream from the AGA in the P-site of the ribosome. At this point, one of at least four things could occur happen (see Figure 20):

A. A pause occurs, but nothing is bound to the A-site. The nascent polypeptide, no longer being extended, falls off and is degraded.

B. The AGA binds a release factor, the polypeptide disassociates from the ribosome, is again non-functional and is subsequently degraded.

C. A tRNA-Arg (UCG) binds to the AGA codon. Though this is an A:G mismatch in the first position, both the second and third positions have Watson-Crick base pairing.

D. The AGA codon binds a tRNA-Ser (GCU), with the same purine:purine mismatch as above in the third position. The ribosomal stall may promote either of these last two events by causing the ribosome to be more amenable to accepting non-cognate tRNAs.

Through the sequencing of the complete *T. scripta* mitochondrial genome, we have discovered a novel programmed frameshift recoding event. A second frameshift recoding event, well documented in the *nad3* gene of a number of turtles and birds, has been examined from a broad range of chelonians. This survey fortuitously revealed a third, unique recoding event seemingly unrelated to frameshifting, whereby an AGA, classically a stop codon in vertebrate mitochondria, is present in-frame. These findings provide insight into recoding, specifically +1 frameshifting and codon redefinition, as well as mitochondrial translation and ribosomal function.

FUTURE WORK

Obtaining *nad3* protein sequence for a number of turtle and bird species would answer many questions with regards to the mechanism of conserved frame maintenance in those genes where extra nucleotides that disrupt the reading frame are present. This would localize where the shift in frames takes place during translation. Beyond that, further experiments may reveal which erroneous protein products are produced, providing insight into the possible pathways that translation can follow at a putative ribosomal pause site.

The case of the Reeve's turtle, where the *nad3* frameshift site evidently results in cognate decoding at the P-site (CUA AGU AGC), also remains an unresolved discrepancy. It is unfortunate that I was unable to obtain a DNA sample of this species in order to sequence this region. Confirming this sequence is also of high priority in order to refine the mechanisms of programmed translational frameshifting in mitochondria. If the sequence does indeed contain a cognate CUA codon for the P-site along with a frameshift insertion, serious reconsideration of the frameshift mechanisms thought to be used in vertebrate mitochondrial genomes will be necessary.

Further sequencing of chelonian mitochondrial genomes would also be useful. Ideally, this will reveal other novel programmed translational frameshift sites and examples of AGR codons used in-frame. Expanding the number of sequences carrying frameshift sites increases the power of statistical and bioinformatics investigations into

frameshift correlates, possibly revealing other important sequences and structures in the mRNA that stimulate frameshifting. The complete *M. parkeri* sequence would be especially interesting, as it would allow the investigation of the ribosomal genes and the proposed tRNA-like release factor structures contained within them.

The ideal future experimental scenario involves producing a chelonian mitochondrial translational system. With such a system in place, mutational analysis of the various frameshift sites and AGR codon redefinitions could be performed. This is the most direct method to investigate the mechanistic properties of mitochondrial programmed translational frameshifting, and would provide a wealth of information on frameshift stimulatory sites and chelonian mitochondrial translation in general.

APPENDIX: PRIMER PAIRS

Complete list of all primer pairs used in this study. All primers are written 5 '-3 '	
Chicken DNA and mRNA primers	
GGDNA1 CTGATGAGGATCTTGCTCTTCT	GGDNA2 TAATTTGCTGGGTCGAAACCT
GGRNA2 AGCAGCCTGATACTGACAC	GGRNA Poly-T TTTTTTTTTTTTTTTTTTTTTTT

***nad3* frameshift region primers**	
SnakeND3F1 GAAGCAGCARTMTGATACTGACAC	SnakeND3R2 CGTRTARTTGCRAYTAYSAGGC
6TND3-P1 CCCCATAYGAGTGYGGATTYGACCC	TND3P2V3 GCTCATTCTAGKCCTCCTTGRATTC

***T. scripta* mitochondrial genome sequencing primers**	
PHE49F ACGGCACTGAAGATGCCAAGATG	12S1021R GATGTTCCAAGTACACCTTCCGG
12S275F AAAATTAAGCAATAAGCATAAGC	16S1957R GTATCTTTTTGGTAAACAGTCGGG
16S1907F AAACTGTTAACCCAACACAGGAGCG	16S2537R CTCCGGTCTGAACTCAGATCACGT
16S2490F GACCTCGATGTTGGATCAGGAC	ND13637R CGTATCGGAATCGTGGATAAGAGGC
ND13080F AGCCCTATCAATCTGACTCCCACT	TND2R395 (4372, R) GGNGCTARTTTTTGTCAGG
ND24167F CCACCAAATATTTTCTAACACAAG	ND24726R GTTAGTGGTGGTAGGCCTGCGAGTG
ND24504F GGATTAAACCAAACCCAACTACG	TCOX1R1 (5474, R) GCTGTRCCDACTATDCCTGCTC
TRP5059F GGATCAAACCTATTAAACCAGAG	CO16281R GTCTATCCCAACGGTAAATATGTGG
CO16064F ATCTAAATACAACCTTCTTCGACC	CO16877R GTGGTGGGCAGCCATGTAGTCATTC
CO16839F CTGAGAAGCATTCTCCTCAAAGCG	A87846R TCATGGGTCTGGATTTAATTGTGGC
CO27557F GAATACTAATCTCAGCTGAAGACG	A68018R CTTATGAATTGGTCGAAGAATG
CO27557F GAATACTAATCTCAGCTGAAGACG	TCOX3R403 (9096, R) GTGATWGTTACNCCTGAGGC

Complete list of all primer pairs used in this study. All primers are written 5 '-3 '	
A88039F CATTCTTCGACCAATTCATAAG	CO39118R TTTCTATTAGACTGTGGTGGGCTC
TATP6F500 (8498, F) GACTNACAGCYAACYTAACAGC	CO39400R CATCTACGAAGTGTCAATATC
CO39034F AGAACTAGGTGGATGTTGACCACC	TND3P2V3(9854) GCTCATTCTAGKCCTCCTTGRATTC
ND39769F GCACTACTCCTACCTCTACCATGAG	ND4R359(10613) TGTTGGGATTAGTGTGGCTTC
ND410586F GTACTTGCCTTCTCAGCCACAG	ND411612R ATATTAACAGCAAAAAGTCCC
ND411088F CTCCATCTGCTTACGACAAACAGA	ND512961R TGAATGGTATTCCTGTGAGTGCTA
ND411543F ACTATCCCACCAACCCACACACGAG	ND512961R TGAATGGTATTCCTGTGAGTGCTA
ND512038F TAACTTTATTGGGCACATCC	ND512961R TGAATGGTATTCCTGTGAGTGCTA
ND512574F CCCAGTCTCAGCATTACTACACTC	ND513582R AGGTTATGTAGATTTTAATTAGGCC
ND513500NF AGCTGATACATCTTGATTCG	CYTB14332R GTAGTGATCCRAAGTTTCATC
ND513500NF AGCTGATACATCTTGATTCG	CYTB14700R GTGTTACCAATGTATGGAATGG
CYTB14416F CTACTCACCAGACATCTCCATAGC	CYTB15108R AAGGATAGAGAGTAGTAGGGC
CYTB15006F ACCCAGATAACTTCACACCAG	UNKNR2 TTATGTCAGGTTAGTTGCCCTCG
CYTB15360F GCAGGTATAATCGAAAACAAAATAC	UNKNR1 CGGATTTAGGGGTTTGACGAGGA
UNKNF1 TAAATCCGAGGGCAACTAACCTG	UNKNR3 TTTGGGCTATCATGGTGTGCCTG
UNK2AF TTTTTAGCTAAACCCCCCTACCCCC	12S74R GTAAGGTTAGGACCAAATCTTTG
UNKNF1 TAAATCCGAGGGCAACTAACCTG	12S118R CACTGGTGTGCTGATACTTGCATGTG

LITERATURE CITED

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25:3389-3402.

Arcari, P. and G.G. Brownlee. 1980. The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria. Nucl. Acids Res. 8:5207-5212.

Baranov, P.V., Gesteland, R.F., and J.F. Atkins. 2002. Recoding: translational bifurcations in gene expression. Gene 286:187-201.

Beckenbach, A.T., Robson, S.K.A., and R.H. Crozier. 2005. Single Nucleotide +1 Frameshifts in an Apparently Functional Mitochondrial Cytochrome b Gene in Ants of the Genus *Polyrhachis*. J. Mol. Evol. 60:141-152.

Belcourt, M.F. and P.J. Farabaugh. 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell 62:339-352.

Blinkowa, A.L. and J.R. Walker. 1990. Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III {gamma} subunit from within the {tau} subunit reading frame. Nucl. Acids Res. 18:1725-1729.

Bruijn, M.H. de and A. Klug. 1983. A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire 'dihydrouridine' loop and stem. EMBO J. 2:1309-1321.

Cai, Q. and K.B. Storey. 1996. Anoxia-induced gene expression in turtle heart. Upregulation of mitochondrial genes for NADH-ubiquinone oxidoreductase subunit 5 and cytochrome c oxidase subunit 1. Eur. J. Biochem.. 241(1):83-92.

Chavatte, L.,Kervestin, S., Favre, A., and O. Jean-Jean. 2003. Stop codon selection in eukaryotic translation termination: comparison of the discriminating potential between human and ciliate eRF1s. EMBO J. 22:1644-1653.

Clare, J. and P.J. Farabaugh. 1985. Nucleotide Sequence of a Yeast Ty Element: Evidence for an Unusual Mechanism of Gene Expression. PNAS 82:2829-2833.

Clary, D.O. and D.R. Wolstenholme. 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Biol. 22:252-271.

Craigen, W.J., Cook, R.G., Tate, W.P., and C.T. Caskey. 1985. Bacterial Peptide Chain Release Factors: Conserved Primary Structure and Possible Frameshift Regulation of Release Factor 2. PNAS 82:3616-3620.

Curran, J.F. 1993. Analysis of effects of tRNA:message stability on frameshift frequency at the *Escherichia coli* RF2 programmed frameshift site. Nucl. Acids Res. 21:1837-1843.

D'Errico, I., Gadaleta, G., and C. Saccone. 2004. Pseudogenes in metazoa: Origin and features. Brief. Funct. Genomic. Proteomic. 3:157-167.

Desjardins, P. and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome : A novel gene order in higher vertebrates. J. Mol. Biol. 212:599-634.

Dörner, M., Altmann, M., Paabo, S., and M. Morl. 2001. Evidence for Import of a Lysyl-tRNA into Marsupial Mitochondria. Mol. Biol. Cell 12:2688-2698.

Farabaugh, P.J. 2000. Translational frameshifting: implications for the mechanism of translational frame maintenance. Prog. Nucl. Acid Res. Mol Biol. 64:131-170.

Farabaugh, P.J. 1996a. Programmed Translational Frameshifting. Annu. Rev. Genet. 30:507-528.

Farabaugh, P.J. 1996b. Programmed translational frameshifting. Microbiol. Rev. 60:103-134.

Farabaugh, P.J., Zhao, H., and A. Vimaladithan. 1993. A novel programed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: Frameshifting without tRNA slippage. Cell 74:93-103.

Flower, A.M. and C.S. Mchenry. 1990. The {gamma} Subunit of DNA Polymerase III Holoenzyme of Escherichia coli is Produced by Ribosomal Frameshifting. PNAS. 87(10): 3713-3717.

Frank, J., Sengupta, J., Gao, H., Li, W., Valle, M., Zavialov, A., and M. Ehrenberg. 2005. The role of tRNA as a molecular spring in decoding, accommodation, and peptidyl transfer. FEBS Letters. 579(4): 959-962.

Gallant, J.A. and D. Lindsley. 1998. Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream. PNAS. 95(23): 13771-13776.

Gesteland, R.F. and J.F. Atkins. 1996. Recoding: Dynamic Reprogramming of Translation. Annu. Rev. of Biochem.. 65(1): 741-768.

Gromadski, K.B. and M.V. Rodnina. 2004. Kinetic Determinants of High-Fidelity tRNA Discrimination on the Ribosome. Mol. Cell. 13(2): 191-200.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Hanada, T., Suzuki, T., Yokogawa, T., Takemoto-Hori, C., Sprinzl, M., and K. Watanabe. 2001. Translation ability of mitochondrial tRNA-Ser with unusual secondary structures in an in vitro translation system of bovine mitochondria. Genes to Cells. 6(12): 1019-1030.

Hansen, T.M., Baranov, P.V., Ivanov, I.P., Gesteland, R.F., and J.F. Atkins. 2003. Maintenance of the correct open reading frame by the ribosome. EMBO Reports. 4(5): 499-504.

Härlid, A., Janke, A., and U. Arnason. 1997. The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. Mol. Biol. Evol. 14(7): 754-761.

Hayashi, I., Kawai, G., and K. Watanabe. 1998. Higher-order structure and thermal instability of bovine mitochondrial tRNASer(UGA) investigated by proton NMR spectroscopy. J. Mol. Biol.. 284(1): 57-69.

Hofacker, I.L., Fekete, M., and P.F. Stadler. 2002. Secondary Structure Prediction for Aligned RNA Sequences. J. Mol. Biol. 319(5): 1059-1066.

Ivanov, I.P., Rohrwasser, A., Terreros, D.A., Gesteland, R.F., and J.F. Atkins. 2000. Discovery of a spermatogenesis stage-specific ornithine decarboxylase antizyme: Antizyme 3. PNAS. 97(9): 4808-4813.

Ivanov, V., Beniaminov, A., Mikheyev, A., and E. Minyat. 2001. A mechanism for stop codon recognition by the ribosome: a bioinformatic approach. RNA. 7(12): 1683-1692.

Jacks, T. and H. Varmus. 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science. 230(4731): 1237-1242.

Kervestin, S., Frolova, L., Kisselev, L., and O. Jean-Jean. 2001. Stop codon recognition in ciliates: Euplotes release factor does not respond to reassigned UGA codon. EMBO Reports. 2(8): 680-684.

King, M.P. and G. Attardi. 1993. Post-transcriptional regulation of the steady-state levels of mitochondrial tRNAs in HeLa cells. J. Biol. Chem. 268(14): 10228-10237.

Klobutcher, L.A. 2005. Sequencing of Random *Euplotes crassus* Macronuclear Genes Supports a High Frequency of +1 Translational Frameshifting. Euk. Cell. 4(12): 2098-2105.

Klobutcher, L.A. and P.J. Farabaugh. 2002. Shifty Ciliates: Frequent Programmed Translational Frameshifting in Euplotids. Cell. 111(6): 763-766.

Kontos, H., Napthine, S., and I. Brierley. 2001. Ribosomal Pausing at a Frameshifter RNA Pseudoknot Is Sensitive to Reading Phase but Shows Little Correlation with Frameshift Efficiency. Mol. Cell. Biol. 21(24): 8657-8670.

Kurland, C.G. 1992a. Evolution of mitochondrial genomes and the genetic code. Bioessays. 14(10): 709-714.

Kurland, C.G. 1992b. Translational Accuracy and the Fitness of Bacteria. Annu. Rev. Genet. 26(1): 29-50.

Larsen, B., Wills, N.M., Gesteland, R.F., and J.F. Atkins. 1994. rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. J. Bacteriol. 176(22): 6842-6851.

Leipuviene, R. and G.R. Bjork. 2005. A reduced level of charged tRNAArgmnm5UCU triggers the wild-type peptidyl-tRNA to frameshift. RNA. 11(5): 796-807.

Li, Z., Stahl, G., and P.J. Farabaugh. 2001. Programmed +1 frameshifting stimulated by complementarity between a downstream mRNA sequence and an error-correcting region of rRNA. RNA. 7(2): 275-284.

Lowe, T.M. and S.R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids Res. 25: 955-964.

Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J., Gesteland, R., and S. Hayashi. 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. Cell. 80(1): 51-60.

McFarland, R., Kirby, D.M., Fowler, K.J., Ohtake, A., Ryan, M.T., Amor, D.J., Fletcher, J.M., Dixon, J.W., Collins, F.A., Turnbull, D.M., Taylor, R.W., and D.R. Thorburn. 2004. De novo mutations in the mitochondrial *ND3* gene as a cause of infantile mitochondrial encephalopathy and complex I deficiency. Ann. Neurol. 55(1): 58-64.

Mellor, J., Fulton, S., Dobson, M., Wilson, W., Kingsman, S., and A. Kingsman. 1985. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. Nature. 313(5999): 243-246.

Milbury, C. and P. Gaffney. 2005. Complete Mitochondrial DNA Sequence of the Eastern Oyster *Crassostrea virginica.* Mar. Biotechnol. 7(6): 697-712.

Mindell, D.P., Sorenson, M.D., and D.E. Dimcheff. 1998. An Extra Nucleotide is not Translated in Mitochondrial ND3 of Some Birds and Turtles. Mol. Biol. Evol. 15(11): 1568-1571.

O'Connor, M. 2002. Imbalance of tRNA-Pro isoacceptors induces +1 frameshifting at near-cognate codons. Nucl. Acids Res. 30:759-765.

Parker, J. 1989. Errors and alternatives in reading the universal genetic code. Microbiol. Rev. 53(3): 273-298.

Pereira, S.L. and A.J. Baker. 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. BMC Evol. Biol. 4(17):

Plant, E.P. and J.D. Dinman. 2005. Torsional restraint: a new twist on frameshifting pseudoknots. Nucl. Acids Res. 33(6): 1825-1833.

Poole, E., Brown, C., and W. Tate. 1995. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. EMBO J. 14(1): 151-158.

Raman, A., Guarraia, C., Taliaferro, D., Stahl, G., and P.J. Farabaugh. 2006. An mRNA sequence derived from a programmed frameshifting signal decreases codon discrimination during translation initiation. RNA. 12(7): 1154-1160.

Rodnina, M.V., Gromadski, K.B., Kothe, U., and H.J. Wieden. 2005. Recognition and selection of tRNA in translation. FEBS Letters. 579(4): 938-942.

Rodnina, M.V., Pape, T., Fricke, R., Kuhn, L., and W. Wintermeyer. 1996. Initial binding of the elongation factor Tu.GTP.aminoacyl-tRNA complex preceding codon recognition on the ribosome. J. Biol. Chem. 271(2): 646-652.

Seligmann, H. and D.D. Pollock. 2004. The Ambush Hypothesis: Hidden Stop Codons Prevent Off-Frame Gene Reading. DNA Cell Biol. 23(10): 701-705.

Shah, A.A., Giddings, M.C., Parvaz, J.B., Gesteland, R.F., Atkins, J.F., and I.P. Ivanov. 2002. Computational identification of putative programmed translational frameshift sites. Bioinformatics. 18(8): 1046-1053.

Siebert, S. and R. Backofen. 2005. MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. Bioinformatics. 21(16): 3352-3359.

Somogyi, P., Jenner, A.J., Brierley, I., and S.C. Inglis. 1993. Ribosomal pausing during translation of an RNA pseudoknot. Mol. Cell Biol. 13(11): 6931-6940.

Spanjaard, R.A., Chen, K., Walker, J.R., and J.V. Duin. 1990. Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg). Nucl. Acids Res. 18(17): 5031-5036.

Spanjaard, R.A. and J.V. Duin. 1988. Translation of the Sequence AGG-AGG Yields 50% Ribosomal Frameshift. PNAS. 85(21): 7967-7971.

Steinberg, S., Gautheret, D. and R. Cedergren. 1994. Fitting the structurally diverse animal mitochondrial tRNAs(Ser) to common three-dimensional constraints. J. Mol. Biol. 236(4): 982-989.

Stewart, J.B. and A.T. Beckenbach. 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome. 48(1): 46-54.

Sundararajan, A., Michaud, W.A., Qian, Q., Stahl, G., and P.J. Farabaugh. 1999. Near-Cognate Peptidyl-tRNAs Promote +1 Programmed Translational Frameshifting in Yeast. Mol. Cell. 4(6): 1005-1015.

Thompson J.D., Higgins D.G., and T.J. Gibson. 1991. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22(22): 4673-4680.

Tsuchihashi, Z. 1991. Translational frameshifting in the Escherichia coli dnaX gene in vitro. Nucl. Acids Res. 19(9): 2457-2462.

Tsuchihashi, Z. and A. Kornberg. 1990. Translational Frameshifting Generates the {gamma} Subunit of DNA Polymerase III Holoenzyme. PNAS. 87(7): 2516-2520.

Tu, C., Tzeng, T., and J.A. Bruenn. 1992. Ribosomal Movement Impeded at a Pseudoknot Required for Frameshifting. PNAS. 89(18): 8636-8640.

Vimaladithan, A. and P.J. Farabaugh. 1994. Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. Mol. Cell Biol. 14(12): 8107-8116.

Watanabe, Y., Kawai, G., Yokogawa, T., Hayashi, N., Kumazawa, Y., Ueda, T., Nishikawa, K., Hirao, I., Miura, K., and K. Watanabe. 1994. Higher-order structure of bovine mitochondrial tRNA(SerUGA): chemical modification and computer modeling. Nucl. Acids Res. 22(24): 5378-5384.

Weiner, A.M. and K. Weber. 1973. A single UGA codon functions as a natural termination signal in the coliphage Q[beta] coat protein cistron. J. Mol. Biol. 80(4): 837-855.

Weiss, R., Dunn, D., Atkins, J., and R. Gesteland. 1987. Slippery runs, shifty stops, backward steps, and forward hops: -2, -1, +1, +2, +5, and +6 ribosomal frameshifting. Cold Spring Harb. Symp. Quant. Biol. 52(687-693.

Weiss, R. and J. Gallant. 1983. Mechanism of ribosome frameshifting during translation of the genetic code. Nature. 302(5907): 389-393.

Wilson, W., Malim, M.H., Mellor, J., Kingsman, A.J., and S.M. Kingsman. 1986. Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. Nucl. Acids Res. 14(17): 7001-7016.

Zardoya, R. and A. Meyer. 1998. Complete mitochondrial genome suggests diapsid affinities of turtles. PNAS. 95(24): 14226-14231.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acids Res. 31(13): 3406-15.