

CODE WITHIN CODES: CODON USAGE REGULATE PROTEIN EXPRESSION,
STRUCTURE, AND FUNCTION

APPROVED BY SUPERVISORY COMMITTEE

Yi Liu, Ph.D., Supervisor

Jin Jiang, Ph.D., Chairperson

Melanie Cobb, Ph.D.

Carla Green, Ph.D.

DEDICATION

To my loving husband, Jiefu Yin, who has provided me with his endless encouragement through the years. He never hesitates to devote his unselfish love through the long physical distance between us. Accomplishing our study and getting together are my unwavering motivations to overcome frustration and achieve my goal during the past five years. For the sacrifices and supports he has made, I am eternally grateful.

To my dear parents and parents-in-law. They unconditionally love us and selflessly support us, especially on our study and our family during Ph.D. process. For their advice, their patience, and their faith, I am greatly grateful.

CODE WITHIN CODES: CODON USAGE REGULATE PROTEIN EXPRESSION,
STRUCTURE, AND FUNCTION

by

JINGJING FU

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2018

Copyright

by

Jingjing Fu, 2018

All Rights Reserved

CODE WITHIN CODES: CODON USAGE REGULATE PROTEIN EXPRESSION,
STRUCTURE, AND FUNCTION

Jingjing Fu

The University of Texas Southwestern Medical Center at Dallas, 2018

Supervising Professor: Yi Liu, Ph.D.

Most amino acids are encoded by two to six synonymous genetic codons. Synonymous codons are not used with the same frequency in all organisms, and every organism has its own preferred codon usage bias. Codon usage bias has been shown to positively correlate with tRNA abundance, thus optimal codons are thought to be translated more efficiently and accurately. Consistent with this, strong codon usage biases have been shown to be important for the expression of highly expressed genes in different organisms, and codon optimization has been widely used to enhance heterologous protein expression. Therefore, codon usage can be an important determinant in gene expression. In addition, codon usage has been shown to influence translation elongation rate and protein structure by affecting the co-translational folding process

in *E. coli*, fungi, and insects. In addition to its role in regulating protein translation, codon usage also has a major role in determining the level of gene expression through transcriptional and post-transcriptional processes. As such, gene codon usage has been proposed to be a code within the genetic code that can determine both gene expression levels and protein structures and therefore activity. However, the effects of codon usage in multi-tissue organisms, for example, animals and humans, are not clear.

In the first part of the thesis, by codon-optimizing open reading frame of *Drosophila* period gene, I showed that dper codon usage is critical for its circadian clock function. Optimization of dper codon usage resulted in conformational changes of dPER protein, altered dPER phosphorylation profile and stability, and impaired dPER repressor function in the circadian negative feedback loop.

In the second part of the thesis, I reported that changing rare codons to common in KRAS increased translation and mRNA levels. Regulation of mRNA levels is a major mechanism affecting KRAS levels, but the effect was not a product of mRNA stability, but instead transcriptional. Moreover, codon usage also had an impact on the structure of KRAS. Thus, the rare codon bias of KRAS effects more aspects of protein production and function than previously appreciated, which has important implications for other rare codon enriched mammalian genes.

TABLE OF CONTENTS

CHAPTER 1. Introduction

Codon Usage Bias	1-5
Drosophila Circadian System as an Animal System to Study Codon Usage Effect	5-8
The Involvement of Codon Usage Regulation in Human Ras Superfamily	8-10

CHAPTER 2. Codon Usage Affects the Structure and Function of the Drosophila

Circadian Clock Protein PERIOD

N-Terminal Codon Optimization of <i>per (op1)</i> Impairs Circadian Behavioral Rhythms ..	16-18
Impaired Molecular Rhythms in OP1 Flies	18-19
Impaired dPER Function in the Circadian Negative Feedback Loop in OP1 Flies	19-20
Wild-type dPER Overexpression Does Not Cause Abnormal Circadian Phenotypes	21
Mechanisms of Codon Influence on Circadian System	21-25
Summary	25-29

CHAPTER 3. Codon Optimality Regulates Ras Expression and Structure

Codon Usage Strongly Affects Expression of <i>KRAS</i>	48-49
Rare Codons Suppress <i>KRAS</i> mRNA Translation.....	49-51
<i>KRAS</i> Codon Optimization Increases mRNA Level but Does Not Affect mRNA Stability	51-52
<i>KRAS</i> Codon Usage Regulates Transcription and Chromatin Modifications	52-54
The Differential Effects of Codon Usage on <i>KRAS</i> Expression in Different Cell Lines	54-55
Codon Optimization Alters <i>KRAS</i> Protein Structure	55-56
Summary	57-60

CHAPTER 4. Material and Methods

Calculation of the CAI and CBI.....	75
Calculation of Protein Structural Disorder Tendency	75
Constructs used in Fly system	75
Generation of Transgenic Flies	75-76
Drosophila Locomotor Activity Analysis.....	76-77
Immunoblotting.....	77-78
Immunohistochemistry	78-79
Quantitative Real-time PCR	79-80
Immunoprecipitation Assays	80
Trypsin Sensitivity Assay	80-81
Thermal Shift Assay	81
S2 Culture and Transfection	81-82
GST Pull-Down Assay	82
Mammalian Cell Lines.....	82-83
Plasmids Used in Mammal system	83
RNA analysis	83-84
Nuclear Run-On Assay	84
Chromatin Immunoprecipitation Assay.....	85
In vitro Transcription	85
In vitro Translation using Mammalian HEK-293T and <i>S. cerevisiae</i> Cell-Free Lysates	86-87
Polysome Profiling	87-88
CHAPTER 5. BIBLIOGRAPHY.....	89-94

PRIOR PUBLICATIONS

1.Fu J, Christopher C, Liu Y. Codon optimality regulates oncogene Kras gene expression.

Submitted. 2018.

2.Fu J, Murphy KA, Zhou M, Li YH, Lam VH, Tabuloc CA, Chiu JC, Liu Y. Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev*

2016, Aug; 30(15); 1761-75.

3.Zhou Z, Dang Y, Zhou M, Li L, Yu CH, Fu J, Chen S, Liu Y. Codon usage is an important determinant of gene expression levels through its effects on transcription. *Proc. Natl. Acad. Sci.*

U.S.A. 2016 Oct 11; 113(41); 6117-25

4.Zhou M, Wang T, Fu J, Xiao G, Liu Y. Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol. Microbiol*, 2015, Sep; 97(5); 974-87.

LIST OF FIGURES

CHAPTER 1. Introduction

Figure 1.1	13
Figure 1.2	14
Figure 1.3	15

CHAPTER 2. Codon Usage Affects the Structure and Function of the *Drosophila* Circadian Clock Protein PERIOD

Figure 2.1	30-31
Figure 2.2	32-33
Figure 2.3	34-35
Figure 2.4	36
Figure 2.5	37-38
Figure 2.6	39-40
Figure 2.7	41
Figure 2.8	42
Figure 2.9	43-44
Figure 2.10	45
Figure 2.11	46

CHAPTER 3. Codon Optimality Regulates Ras Expression and Structure

Figure 3.1	61
Figure 3.2	62
Figure 3.3	63

Figure 3.4	64
Figure 3.5	65
Figure 3.6	66
Figure 3.7	67
Figure 3.8	68
Figure 3.9	69
Figure 3.10	70-71
Figure 3.11	72

LIST OF TABLES

CHAPTER 1. Introduction

Table 1.1	11
Table 1.2	12

CHAPTER 2. Codon Usage Affects the Structure and Function of the *Drosophila* Circadian

Clock Protein PERIOD

Table 2.1	47
-----------------	----

CHAPTER 3. Codon Optimality Regulates Ras Expression and Structure

Figure 3.1	74
------------------	----

LIST OF ABBREVIATIONS

CAI	Codon adaptation Index
tAI	tRNA adaptation Index
E. coli	Escherichia coli
frq	frequency
FRQ	FREQUENCY
dPER	Drosophila PERIOD
PTT	Premature Transcription Termination
Adh	Alcohol dehydrogenase
SNP	Signal Nucleotide Polymorphism
hPER	Homo. Sapiens PERIOD
MDR1	Multiple-Drug Resistance 1
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
dCLK	Drosophila CLOCK
dTIM	Drosophila TIMELESS
DBT	DOUBLETIME
RAS	Rat Sarcoma
HRAS	Harvey Rat Sarcoma Viral Oncogene Homolog
KRAS	Kristen Rat Sarcoma Viral Oncogene Homolog
NRAS	Neuroblastoma Rat Sarcoma Viral Oncogene Homolog
GEF	Guanine nucleotide exchange factor
GAP	GTPase activating protein

MAP3K	MAP kinase kinase kinase
MEK	MAP kinase kinase
ERK	MAP kinase
per0	per-null
WT	Wild Type
LD	12h:12h Light: Dark
DD	Constant Darkness
ZT	Zeitgeber Time
CT	Circadian Time
PDF	Pigment Dispersing Factor
IHC	Immunohistochemistry
dcwo	Drosophila clockwork orange
dgol	Drosophila goliath
S2	Drosophila Schneider cells
SLIMB	SCF ubiquitin ligase complex
GST	Glutathione S-Transferase
CCID	CLK/CYC inhibition domain
IDP	Intrinsically Disordered Protein
ALS	Amyotrophic Lateral Sclerosis
MDR1	Multidrug Resistance 1
CTD	Carboxyl-Terminal Domain
ChIP	Chromatin Immunoprecipitation
H3K4	Histone H3 lysine 4

H3K9	Histone H3 lysine 9
H3K4me3	Histone H3 lysine 4 trimethylation
H3K9ac	Histone H3 lysine 9 acetylation
HAT	Histone Acetyltransferase
EDTA	EthyleneDiamineTetraacetic Acid
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonylfluoride
qRT-PCR	real-time PCR
RT	Room Temperature
SDS	Sodium Dodecyl Sulphate
CHX	Cycloheximide
RNasin	RNase inhibitor

ACKNOWLEDGEMENT

It is my pleasure to acknowledge the roles of the individuals who offered generous supports for the completion of my Ph.D. research:

I would like to express my sincere gratitude to my Ph.D. advisors, Dr. Yi, Liu, who taught me the art of scientific thinking, trained me on physiology-related fields, and guided me to conduct the research projects. His enthusiasm on scientific research always encourages me to be a scientist. I truly enjoyed working in a research environment that stimulates original thinking.

I would like to thank my committee members, Drs. Jin Jiang, Melanie Cobb and Carla Green, for advice, feedback, and encouragement over the years.

I have had the great opportunity to work in collaboration with many wonderful and collaborative postdocs and graduate students in our lab. I truly appreciate their guidance on new techniques and insightful discussions regarding my research projects. I am also extremely fortunate to have collaborated with Drs. Joanna Chiu from UC Davis, and Chris Counter from Duke University. I would like to thank many faculty members at UTSW, including Drs. Robin Hiesingers, Adeian Rothenfluh, Huaqi Jiang, and Chengcheng Zhang for their help..

This work would not materialize without the financial support from National Institutes of Health and Welch Foundation (I-1560).

I would also like to acknowledge Department of Physiology, for the support on lab resources and technical support. I am appreciative of the guidance and help from wonderful Integrative Biology Program, including program chair Dr. Yi Liu, program coordinator Priyarama Sen. Finally, I want to thank the Division of Biological Science Graduate program at UT Southwestern Medical Center.

CHAPTER ONE

INTRODUCTION

Codon usage bias

Codon usage bias

Most amino acids are encoded by two to six synonymous codons. Preferential use of certain synonymous codons, a phenomenon called codon usage bias, was found in all genomes (Ikemura 1985; Sharp et al. 1986; Comeron 2004; Plotkin and Kudla 2011). The mutation-selection-drift balance model is the general theory commonly used to explain the existence of codon usage bias (Sharp and Li 1986; Bulmer 1987; Hershberg and Petrov 2008). The balance between forces of mutational and selective mechanisms generate the unequal frequencies of synonymous codons within a finite population. Specifically, mutational biases uniformly affect the GC-content of the whole genome, whereas selective forces act on coding sequences in a gene specific manner. The theory emphasizes the cause of selection on codon usage bias is generally attributed to translational forces, as codons corresponding to more abundant cognate tRNAs are translated more efficiently and accurately by reducing ribosome pausing during translational elongation process and decreasing the probability of incorporating incorrect amino acids (Ikemura 1981 {Stoletzki, 2007 #44; Gingold and Pilpel 2011}). Besides the consensus roles of codon usage bias in translational process, emerging evidence has shown the diverse effects of codon usage bias in different layers of gene regulation, which will be discussed in detail bellowed.

In Prokaryotes, codon usage has been shown to play a role in endogenous gene

expression, as codons in highly expressed genes correlate with highly abundant tRNAs (Ikemura 1985). By studying the heterologous protein expression in *E. coli*, it was suggested that translation rate and synonymous codon usage can affect protein expression, folding and functions (Komar et al. 1999; Zhang et al. 2009; Siller et al. 2010; Spencer et al. 2012). A study based on the analysis of tRNA adaptation index (tAI) along *E. coli* genes further showed that codon usage could mediate ribosome dynamics by acting in the very beginning 5' region of the coding sequence (Tuller et al. 2010). tAI is a measure of the translational efficiency of a particular codon after taking the imperfect pair between tRNAs and wobble position of codons into account. Lower tAI value in the first 20 codons was proposed to serve as a translational 'ramp' to reduce ribosomal traffic jams. Moreover, the ribosomes recognizing the suboptimal codons within these 'ramps' may be sensitive to nutrient abundance in cells, which could prevent more translation of the message when conditions are not ideal. This phenomenon is also conserved in other species, such as yeast and humans.

More and more studies have demonstrated the importance of codon usage in gene regulation in eukaryotes. As in eukaryotic organisms, transcription and translation processes are separated both in time and space. This is an essential difference with prokaryotes, in which translation occurs just during transcription elongation process. Codon usage bias, therefore, shows more differential effects in eukaryotes. In yeast, A/T biased codons are more preferred to be used, and skewed correlation has been shown between codon usage and tRNA abundance (measured by tRNA gene copy number) (Ikemura 1985; Akashi 2003). This is also consistent with the correlation between gene expression levels with codon usage bias in the fission yeast (Hiraoka et al. 2009). Computational analysis using yeast genome-wide data shows that codon

bias is linked to mRNA folding structure (Trotta 2013), and nonoptimal codons in an mRNA can destabilize mRNA during protein translation (Presnyak et al. 2015).

The filamentous fungus *Neurospora crassa* has been shown as an important model organism for codon usage bias study. It exhibits a strong codon usage bias for C/G at wobble codon positions, which is opposite from the codon preference profiles of yeast. Our lab previously showed that the codon usage bias in the *Neurospora* circadian clock gene, frequency (*frq*), is critical for the structure and function of FRQ in vivo (Zhou et al. 2013). More recently, our lab demonstrated that codon usage regulates the speed of mRNA translation elongation and by doing so, affecting co-translational protein folding in *Neurospora* (Yu et al. 2015). Bioinformatic analyses have uncovered correlations between codon usage and potential to form certain protein structural motifs (Zhou et al. 2009; Pechmann and Frydman 2013; Pechmann et al. 2014; Zhou et al. 2015). Codon usage has also been shown to be an important determinant of gene transcription levels (Zhou et al. 2016). However, unlike in yeast, this effect does not work through mRNA stability, whereas by influencing chromatin structure instead. More recently, our lab demonstrated that, rare codons could lead to the formation of putative poly(A) signals, result in premature transcription termination (PTT) within open reading frames, and cause the abolishment of full-length mRNA (Zhou et al. 2018). This mechanism shows a co-evolution between codon usage bias and transcription termination machinery to repress premature termination of transcription and allow for optimal gene expression. Together, these studies in fungi system demonstrated that codon usage can regulation gene expression beyond translation process.

However, there is still some weakness using unicellular organism, as selection on codon

bias is not affected by the different expression profiles of differentiation cells. Current progress on codon usage study mainly comes from bioinformatics analysis and experimental evidence in microorganisms. However, biological functions of codon usage, especially in animal systems, are not clear. Moreover, bioinformatics analysis has suggested an important role of codon usage pattern in the differentiation and regulation of tissue-specific gene products (Plotkin et al. 2004), it is also lack of experimental evidence to prove this observation.

I excitingly found that *Drosophila* is an invaluable model for codon usage study in animals. The nonrandom usage of synonymous codons is a well-established phenomenon in *Drosophila* (Hambuch and Parsch 2005). In particular, among 12 *Drosophila* species, translational selection strength on codon usage is highest in *D. melanogaster* (Heger and Ponting 2007). This strong selection has been predicted to enhance the accuracy of protein synthesis, especially in the highly expressed genes.

Bioinformatics analyses indicate that *Drosophila* genome has a strong codon bias for G/C at the wobble positions (Table 1.1) (Kanaya et al. 2001). The average of CBI of 0.23 for *Drosophila* genes is similar to that of *Neurospora* (0.22) (Zhou et al. 2013), indicating a strong codon usage bias in *Drosophila*. A positive correlation has also been shown between codon usage and gene expression levels in flies (Duret and Mouchiroud 1999). A primary but promising study has stated that codon alteration in alcohol dehydrogenase (*Adh*) gene can change liability of flies after alcohol treatment (Carlini and Stephan 2003).

In mammals, the effect of codon usage is still under debate. This is mainly caused by an evolutionary argument against selective pressure, known as the nearly neutral theory (Sharp et al. 1995; Chamary et al. 2006; Parmley et al. 2006). It is commonly accepted that a mutation will

only be selected for/against if the effect of the mutation on organism fitness multiplied by the effective population size is large. With mammals, however, both the effective population size and the expected effect on organism fitness of a synonymous mutation are quite small, therefore, the mutation will not be strongly selected. It has been long suggested that there may not be the signature of selective pressure on mammalian codon usage {Chamary, 2006 #48}.

In contrast with this consensus views, Results from recent studies have revealed the role of codon usage in mammals, which could be evidenced by the biased codon usage in Homo Sapiens (Table 1.2), and some existence of synonymous SNPs within the coding regions of several human genes. For example, patients with T2434C synonymous polymorphism in exon 18 of hPER1 perform extreme diurnal performance (Carpen et al. 2006), and other group with G2114A in hPER2 were also linked to diurnal preference (Matsuo et al. 2007). However, the mechanism is not clear. Synonymous SNPs within the coding region of the human transporter protein MDR1 alter its protein folding and subsequent transport function (Kimchi-Sarfaty et al. 2007). Moreover, single synonymous SNP in the human Cystic fibrosis transmembrane conductance regulator (CFTR) gene results in Cystic fibrosis, probably due to an alternation of protein stability (Bartoszewski et al. 2010) and conformation (Lazrak et al. 2013b). The codons used by an mRNA could direct a post-translational modification of the proteins (Zhang et al. 2010). In addition, codon optimization can increase the expression of heterologous genes ectopically introduced into mammalian cells (Zolotukhin et al. 1996; Fath et al. 2011).

Taken together, codon usage bias is a critical mechanism in gene regulation, from prokaryotes to eukaryotes.

Drosophila Circadian System as an Animal System to Study Codon Usage Effect

In our previous studies of *Neurospora* circadian clock, our lab discovered that non-optimal codon usage in the frequency (*frq*) gene determines not only FRQ protein expression but also its structure and function (Zhou et al. 2013). In addition, our lab has shown that the non-optimal codon usage in *Kai* genes is important for circadian clock conditionality in the Cyanobacterium *Synechococcus* (Xu et al. 2013). These results established two of first in vivo examples of how non-optimal codon usage in both eukaryotic and prokaryotic organisms regulates protein expression, structure, and function. Because of the important implications of our discoveries and the existence of codon usage bias in almost all organisms, i will use circadian clock proteins as entry points and combine genetic, biochemical, and bioinformatic approaches to study the role and mechanism of codon usage in determining protein structure and function in animal systems.

Circadian clocks control daily oscillations in numerous biological processes and increase fitness of various organisms (Ouyang et al. 1998; Johnson 2001; Ule et al. 2003; Dong et al. 2010; Xue et al. 2012). The importance of circadian clocks is reflected by wide-spread circadian control of gene expression, development, behavior, and physiological activities from fungi to man (Dunlap 1999; Young and Kay 2001; Sehgal 2004; Bell-Pedersen et al. 2005). In human, circadian clocks have ubiquitous influence on a wide range of processes, including sleep/wake and body temperature cycles (Johnson et al. 1981; Moore-Ede et al. 1982; Winfree 1982; Hastings et al. 1991), endocrine functions (Brownstein and Axelrod 1974), drug resistance (Halberg 1982), and the phenomenon of jet lag. The core eukaryotic circadian oscillators consist

of transcription- and translation-based negative feedback loops. Despite evolutionary distance, the circadian oscillator mechanisms from fungi to animals share remarkable similarities and conservation (Liu and Bell-Pedersen 2006; Heintzen and Liu 2007; Baker et al. 2012). I used the *Drosophila melanogaster* system to determine whether codon usage is a universal mechanism for protein structure and function determination in eukaryotes. Similar to *Neurospora*, *D. melanogaster* genome has a strong codon bias for G/C at the wobble positions with an average gene codon bias index of 0.23 (Kanaya et al. 2001; Hambuch and Parsch 2005; Heger and Ponting 2007; Zhou et al. 2013) (Zhou et al. 2015). A positive correlation between codon usage and gene expression levels was previously observed in flies (Duret and Mouchiroud 1999). In addition, introduction of unpreferred codons in the alcohol dehydrogenase (*Adh*) gene in vivo led to reduced protein levels (Carlini and Stephan 2003). As in our studies in *Neurospora*, I reasoned that the sensitivity and robustness of the *Drosophila* circadian system would allow us to uncover the role of codon usage in vivo. In the *Drosophila* circadian system, the CLOCK/CYCLE (CLK/CYC) complex is the positive element in the core circadian negative feedback loop that activates the transcription of period (*dper*) and other clock-controlled genes (Young and Kay 2001; Allada and Chung 2010; Hardin and Panda 2013; Tataroglu and Emery 2015) (Figure 1.1). On the other hand, *dPER* functions as a negative element with TIMELESS (TIM) in the circadian negative feedback loop by interacting and repressing the activity of CLK/CYC complex to inhibit the transcription of *dper* and other CLK/CYC target genes. After its synthesis, *dPER* is progressively phosphorylated by DOUBLETIME (DBT) and other kinases (Kloss et al. 1998; Price et al. 1998; Preuss et al. 2004; Bae and Edery 2006; Gallego and Virshup 2007; Ko et al. 2010; Chiu et al. 2011). A major function of DBT-mediated *dPER* phosphorylation is to trigger

binding to SLIMB (~~α~~-TrCP in mammals), which promotes dPER ubiquitination and subsequent degradation by the proteasome pathway (Grima et al. 2002; Ko et al. 2002; Ko and Edery 2005). After dPER degradation, CLK/CYC is then released from inhibition to start the next round of transcriptional activation of dper and other target genes, resulting in rhythmic transcription.

The Involvement of Codon Usage Regulation in Human Ras Superfamily

RAS family, comprised of three foundation members KRAS, HRAS, and NRAS, are oncogenically mutated in 30% of all cancer cases (Colicelli 2004; Cox and Der 2010; Pylayeva-Gupta et al. 2011; Prior et al. 2012; Cox et al. 2014). They are membrane-bound small GTPases, and act as molecular switches downstream of cell-surface receptors (Figure 1.2). RAS can be activated by growth factor receptors with tyrosine kinase activity (Malumbres and Barbacid 2003). After phosphorylation at tyrosine on these activated receptors, they can serve as docking sites for adaptor proteins with SH2 domains. These adaptor proteins (i.e., GRB2) recruit RAS GEFs (i.e., SOS) from the cytosol to the plasma membrane, which covalently attach to RAS proteins and activate them. One of the classical downstream RAS signaling pathways is RAF kinase cascade (Shields et al. 2000; Shapiro 2002). In these cascades, RAS activates RAF serine/threonine kinases, recruits them to the plasma membrane. RAF is a MAP kinase kinase (MAPKKK), which phosphorylates and activates a MAP kinase kinase (also called MEK). MEK then phosphorylates MAP kinases (also called ERK). Activated ERK translocates into the nucleus, then phosphorylates and activates mitogenic transcription factors.

Over 150 monomeric G proteins can be identified by structural motifs within RAS superfamily (Wennerberg et al. 2005). This superfamily can be classified into five groups based

on similarity in amino acid sequences and protein functions: the RAS, RHO, RAB, RAN and ARF families. In this part, I focused on three foundation members in the RAS family: HRAS, KRAS, and NRAS. Three genes encode almost identical isoforms that are ubiquitously expressed, but are not functionally redundant. The network responses associated with each isoform and individual oncogenic mutations remain to be fully characterized. The encoded proteins share ~85% homology, with the primary differences lying in the last ~23 amino acids (Figure 1.3) (Barbacid 1987). Despite this high homology, multiple lines of evidence suggest that KRAS is expressed at low levels, and moreover, that this low expression is biologically critical to the function of the gene. Namely, KRAS mRNA levels are typically the lowest of the RAS isoforms in human tissues (Fiorucci and Hall 1988) (and see below). Further, increasing the expression of the endogenous murine *Kras* gene results in hyperproliferation of their hematopoietic stem cells and renders the mice more resistant to a carcinogen that induces *Kras*-mutant lung tumors (Pershing et al. 2015; Sasine et al. in press). The importance of the low expression of KRAS cannot be understated, as this gene is essential (Koera et al. 1997), unlike NRAS and HRAS, and is mutated in one quarter of all human cancers, far more than the other two RAS genes (Prior et al. 2012). For example, KRAS is strongly associated with pancreatic, colorectal and lung cancers, whereas NRAS is the isoform most frequently mutated in haemopoietic tumours. HRAS mutations are rarely detected in tumours. In addition, overexpression or amplification of KRAS, but not mutations in the coding sequence, is associated with certain types of cancer (Birkeland et al. 2012; Valtorta et al. 2013). As such, elucidating the mechanism by which KRAS expression is kept low is critical to our understanding of normal and cancer biology. Moreover, I did a research on kinds of mutations of

Kras based on an online somatic mutation database in cancer

(<http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=KRAS>). This database consists of novel mutation sites that have been proven to drive cancers in clinical cases. I found that quite an amount of synonymous mutations have been reported in clinical cases. However, it lacks of experimental evidence to address the pathogenic mechanisms of these synonymous mutations.

Amino Acid	Coden	%	Amino Acid	Coden	%	
Phe	UUU	0.37	Ser	UCU	0.08	
	UUC	0.63		UCC	0.24	
Leu	UUA	0.05		UCA	0.09	
	UUG	0.18		UCG	0.2	
	CUU	0.1		Pro	CCU	0.13
	CUC	0.15			CCC	0.33
	CUA	0.09	CCA		0.25	
	CUG	0.43	CCG		0.29	
Ile	AUU	0.34	Thr	ACU	0.17	
	AUC	0.47		ACC	0.38	
	AUA	0.19		ACA	0.19	
Met	AUG	1		ACG	0.26	
Val	GUU	0.18	Ala	GCU	0.19	
	GUC	0.24		GCC	0.45	
	GUA	0.11		GCA	0.17	
	GUG	0.47		GCG	0.19	
Tyr	UAU	0.37	Cys	UGU	0.29	
	UAC	0.63		UGC	0.67	
Stop codon	UAA	-	Stop codon	UGA	-	
	UAG	-	Trp	UGG	1	
His	CAU	0.4	Arg	CGU	0.16	
	CAC	0.6		CGC	0.33	
Gln	CAA	0.3		CGA	0.15	
	CAG	0.7		CGG	0.15	
Asn	AAU	0.44	Ser	AGU	0.14	
	AAC	0.56	AGC	0.25		
Lys	AAA	0.29	Arg	AGA	0.09	
	AAG	0.71		AGG	0.11	
Asp	GAU	0.53	Gly	GGU	0.21	
	GAC	0.47		GGC	0.43	
Glu	GAA	0.33		GGA	0.29	
	GAG	0.67		GGG	0.07	

Table 1.1 *Melanogaster. Drosophila* codon usage table

Amino Acid	Coden	%	Amino Acid	Coden	%
Phe	UUU	0.46	Ser	UCU	0.19
	UUC	0.54		UCC	0.22
Leu	UUA	0.08		UCA	0.15
	UUG	0.13		UCG	0.05
	CUU	0.13	Pro	CCU	0.29
	CUC	0.2		CCC	0.32
	CUA	0.07		CCA	0.28
	CUG	0.4		CCG	0.11
Ile	AUU	0.36	Thr	ACU	0.25
	AUC	0.47		ACC	0.36
	AUA	0.17		ACA	0.28
Met	AUG	1		ACG	0.11
Val	GUU	0.18	Ala	GCU	0.27
	GUC	0.24		GCC	0.4
	GUA	0.12		GCA	0.23
	GUG	0.46		GCG	0.11
Tyr	UAU	0.44	Cys	UGU	0.46
	UAC	0.56		UGC	0.54
Stop codon	UAA	-	Stop codon	UGA	-
	UAG	-	Trp	UGG	1
His	CAU	0.42	Arg	CGU	0.08
	CAC	0.58		CGC	0.18
Gln	CAA	0.26		CGA	0.11
	CAG	0.74		CGG	0.2
Asn	AAU	0.47	Ser	AGU	0.15
	AAC	0.53		AGC	0.24
Lys	AAA	0.43	Arg	AGA	0.22
	AAG	0.57		AGG	0.21
Asp	GAU	0.46	Gly	GGU	0.16
	GAC	0.54		GGC	0.34
Glu	GAA	0.42		GGA	0.25
	GAG	0.58		GGG	0.25

Table 1.2 *Homo Sapiens* codon usage table

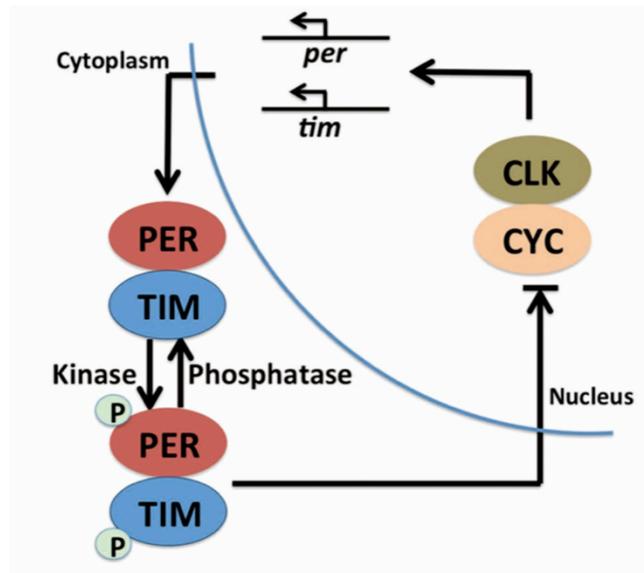


Figure 1.1: Schematic illustrating the negative feedback loop driving the *Drosophila melanogaster*.

CLOCK (CLK) and CYCLE (CYC) activates transcription of period (*per*) and timeless (*tim*). Once these transcripts are expressed and translated, PER and TIM proteins are modified posttranslationally by kinases including DOUBLETIME (DBT), translocate to the nucleus, and repress the activity of CLK and CYC to promote transcription of *per* and *tim*, and other clock-controlled genes. The repression is relieved once PER is phosphorylated at the phosphodegron, including Serine 47, and degraded through the proteasome pathway.

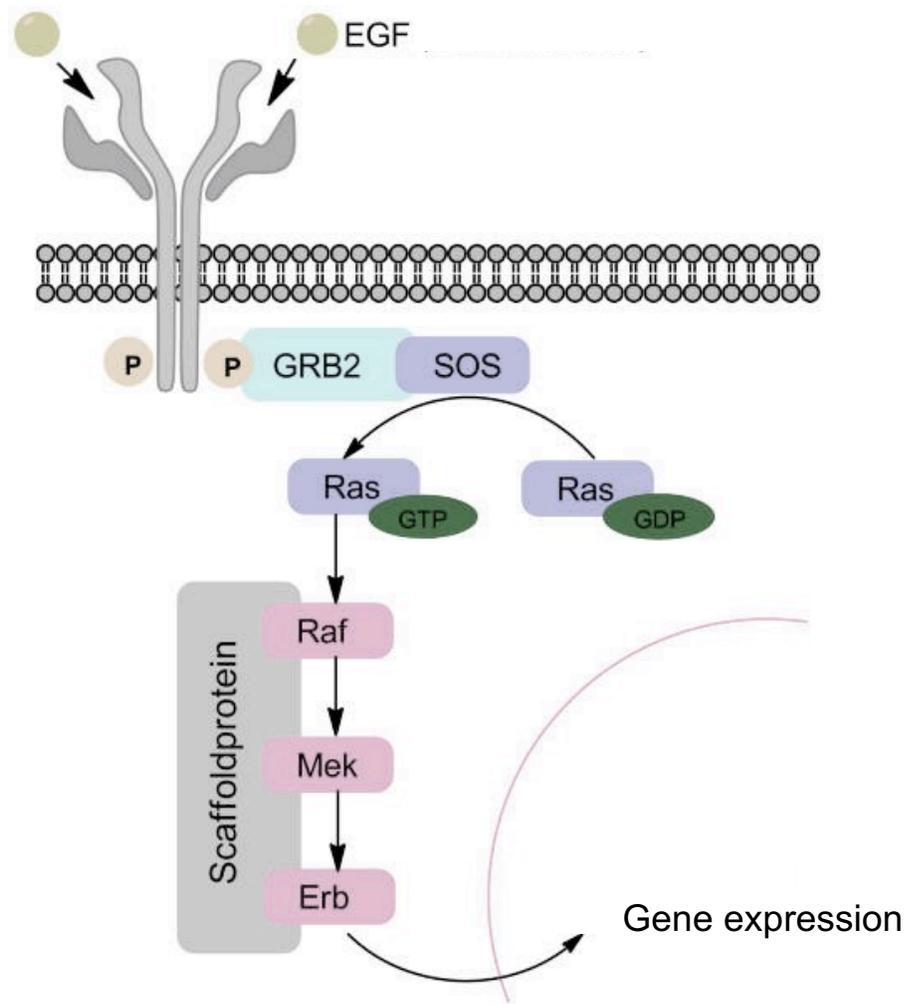


Figure 1.2: RAS signaling pathway.

RAS can be activated by growth factor receptors with tyrosine kinase activity. After phosphorylation at tyrosine on these activated receptors, they can serve as docking sites for adaptor proteins with SH2 domains. These adaptor proteins (i.e., GRB2) recruit RAS GEFs (i.e., SOS) from the cytosol to the plasma membrane, which covalently attach to RAS proteins and activate them. One of the classical downstream RAS signaling pathways is RAF kinase cascade. In these cascades, RAS activates RAF serine/threonine kinases, recruits them to the plasma membrane. RAF is a MAP kinase kinase kinase (MAP3K), which phosphorylates and activates a MAP kinase kinase (also called MEK). MEK then phosphorylates MAP kinases (also called ERK). Activated ERK translocates into the nucleus, then phosphorylates and activates mitogenic transcription factors.

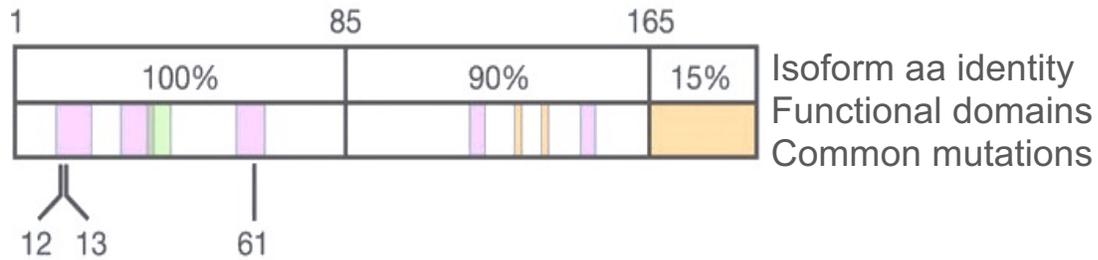


Figure 1.3: Sequence characteristics of human *HRAS*, *KRAS*, and *NRAS*.

These three family members are 80% identical to one another with the N-terminal 85 amino acids completely identical; *HRAS* is 189 amino acids long while *NRAS* and *KRAS* are 188 amino acids in length. The C-terminal 20 amino acid hypervariable domain, responsible for RAS subcellular localization. Amino acid positions 12, 13, 61 are the most common mutations found in RAS-related cancers.

CHAPTER TWO

CODON USAGE AFFECTS THE STRUCTURE AND FUNCTION OF THE DROSOPHILA CIRADIAN CLOCK PROTEIN PERIOD

N-Terminal Codon Optimization of *per* Impairs Circadian behavioral Rhythms

To analyze the codon usage of the *D. melanogaster per* gene (*dper*), I calculated codon usage scores based on the codon adaptation index (CAI) through its open reading frame (ORF) (Figure 1A, Supplemental Fig. S2) (Sharp and Li 1987). In addition, I used the protein secondary structure prediction program IUPRED to predict the locations of intrinsically unstructured/disordered residues of dPER protein. Similar to the *Neurospora* FRQ protein (Zhou et al. 2013; Zhou et al. 2015), most of dPER protein, with the exception of the two known PAS (PERIOD-ARNT-SIM) domains (Huang et al. 1993), were predicted to be intrinsically disordered. This prediction is consistent with the fact that the PAS domains are the only regions of dPER with known crystal structure (Hennig et al. 2009; Merbitz-Zahradnik and Wolf 2015). Interestingly, the predicted intrinsically disordered regions in the N-terminal and middle part of dPER have domains with relatively low codon usage scores. This is consistent with our previous observation that predicted intrinsically disordered regions preferentially use non-optimal codons (Zhou et al. 2015), raising the possibility of a role for codon usage in co-translational protein folding in these regions.

To test this hypothesis, I used a commonly used *dper* rescue vector that contains a 13.2-kb *dper* genomic fragment tagged with the HA epitope at the C-terminus (pCasPeR-*per*-13.2)

that is capable of rescuing *per*⁰ mutant (Lee et al. 1998) and codon optimized the gene region that encodes for the N-terminal end of the dPER protein, including the N-terminal disordered region and part of the PAS A domain (Figure 2.1A). To avoid potential effects on translation initiation and mRNA splicing, the first 10 codons of *dper* and codons near an intron in this region were not changed. The codon-optimized vector was termed p{*dper*(OP1)}, which has the same 5' and 3' regulatory sequences and encodes the identical amino acid sequence as the wild-type gene. Independent lines of transgenic flies harboring the p{*dper*(OP1)} construct were obtained and were evaluated in a *per*-null (*per*⁰) mutant background, in which the circadian clock function could be rescued only when functional dPER is expressed from the transgene. Transgenic flies with the wild-type *dper* construct (p{*dper*(WT)}) were used as control.

To determine if p{*dper*(OP1)} can rescue the circadian clock of *per*⁰ flies, I analyzed the locomotor activity rhythm of the transgenic flies at 25°C. Flies were entrained for 4 days in 12h:12h light: dark (LD) cycles before shifted into constant darkness (DD) to detect free-running rhythms (Figures 2.1B-C, Figure 2.2, Table 2.1). As shown previously, p{*dper*(WT)} flies exhibited normal circadian phenotypes, as indicated by the typical bimodal morning and evening locomotor activity peaks in LD and the continued rhythmicity under constant darkness (Rosato and Kyriacou 2006; Lear et al. 2009), indicating the full rescue of clock function by the wild-type *dper* transgene in the *per*⁰ background. In the heterozygous p{*dper*(OP1)} flies, however, although I observed free-running rhythms in DD, their amplitudes were not as robust as that of the p{*dper*(WT)} flies. The education graphs in Figure 1C showed that the heterozygous p{*dper*(OP1)} flies exhibited significantly reduced anticipation of the morning and evening peaks as indicated by lower anticipation indexes (Cusumano et al. 2009). This indicates that one

copy of p{*dper*(OP1)} transgene is not able to rescue *per*⁰ mutant. Two copies of p{*dper*(OP1)} transgenes also failed to rescue circadian phenotype of the *per*⁰ flies. In fact, p{*dper*(OP1)} homozygous flies exhibited behavioral rhythms that were mostly abolished or severely damped (Figure 1B). In addition, morning and evening peaks in LD were dramatically reduced in the mutant homozygous flies. It should be noted that that much lower than expected number of homozygous p{*dper*(OP1)} flies could be obtained, suggesting that the high dosage of optimized PER protein might have a negative impact on survival. Due to this reason, I used heterozygous p{*dper*(OP1)} flies for the rest of the study to dissect the molecular basis of the observed phenotype. Together, these results indicate that *dper* codon usage is important for its clock function.

Impaired Molecular Rhythms in OP1 Flies

To examine oscillator function at the molecular level, fly head extracts from the heterozygous p{*dper*(WT)} and p{*dper*(OP1)} flies were prepared every 4 h in LD after three full days of entrainment and on the second day in DD. Although p{*dper*(OP1)} showed a cycling of PER protein abundance in LD, I observed a phase advance of about four hours for dPER rhythm, which peaked at ZT16 instead of ZT 20 for the p{*dper*(WT)} flies (Figure 2.3A). In addition, dPER protein levels were also elevated in the mutant flies, suggesting a role for codon optimization in enhancing dPER expression (Figure 2.3B). Consistent with previous studies, there was a robust rhythm of dPER phosphorylation profiles in the p{*dper*(WT)} flies in LD, as indicated by the dPER mobility changes at different time points (dPER mostly hyperphosphorylated at DD24). In contrast, such dPER mobility changes in LD were largely

absent in the $p\{dper(OP1)\}$ flies. The impaired dPER phosphorylation rhythm was also obvious in DD in the mutant flies (Figure 2.3A, bottom panel). These results indicate that dPER molecular rhythms are also impaired in OP1 flies.

Because most dPER protein signals of fly head extracts came from eyes, I examined dPER levels and its cellular localization in the circadian-behavior-related Pigment Dispersing Factor (PDF)-positive neurons in the brain by immunohistochemistry (IHC) (Li et al. 2014). As shown in Figure 2.3C, PDF mostly resided in the cytoplasm. For the $p\{dper(WT)\}$ flies, dPER level was low and was mostly cytoplasmic at ZT16 and became mostly nuclear localized at ZT 20 and 22. Even though a similar temporal change in dPER nuclear localization was also observed in the $p\{dper(OP1)\}$ flies, the levels of $p\{dper(OP1)\}$ were higher at all three time points, reflecting what I observed by Western blots (Figure 2.3A). The greatest difference in dPER signal levels were observed at ZT 16, consistent with an advanced phase of dPER rhythm at the protein levels in the $p\{dper(OP1)\}$ flies (Figure 2.3A). In contrast, no significant differences in PDF levels were apparent between the two fly strains, suggesting that the impaired circadian behaviors in the $p\{dper(OP1)\}$ flies are not due to the changes of PDF (Stoleru et al. 2004).

Impaired dPER Function in the Circadian Negative Feedback Loop in OP1 Flies

Drosophila PER functions as the core negative element in the fly circadian negative feedback loop by repressing the activity of CLK/CYC complex, which results in the transcriptional repression of *dper* and other CLK-CYC target genes. Thus, the elevation of dPER expression in the $p\{dper(OP1)\}$ flies should result in decreased transcription of CLK-CYC target

genes if dPER function in the negative feedback loop is normal. However, the mRNA levels of *dper* and three other direct CLK-CYC target genes (*dtim*, *dcwo* and *dgol*) (Abruzzi et al. 2011) were all significantly elevated in the p{*dper*(OP1)} flies at all time points in LD as compared to the p{*dper*(WT)} flies (Figure 2.4). These results suggest that dPER function in the circadian negative feedback loop was impaired in the p{*dper*(OP1)} flies despite of the increase in dPER levels.

Drosophila PER exerts its role in the negative feedback loop by directly interacting with the CLK/CYC complex and sequestering CLK/CYC off the E-box element from the target gene promoters (Lee et al. 1999; Taylor and Hardin 2008; Menet et al. 2010). Therefore, I compared the interaction between PER and CLK in the p{*dper*(WT)} and p{*dper*(OP1)} flies by immunoprecipitation. The interaction between dPER and CLK are low during mid-day and high near the end of night (Menet et al. 2010). CLK is the limiting factor in the CLK-PER interaction (Bae et al. 2000). CLK immunoprecipitation also showed that there was less PER associated with CLK in the p{*dper*(OP1)} flies (Figure 2.5A). As shown in Figure 2.5B, despite a higher level of dPER in the p{*dper*(OP1)} flies, the amount of CLK associated with dPER was significantly decreased after PER immunoprecipitation. Together, these results suggest that codon optimization resulted in reduced dPER-CLK interaction, providing an explanation for the impaired circadian negative feedback loop in the p{*dper*(OP1)} flies.

I also examined expression profiles of TIM in p{*dper*(OP1)} flies in LD (Figure 2.5C). Consistent with the dPER rhythm, TIM rhythm was also phase advanced in the p{*dper*(OP1)} flies with elevated TIM levels.

Wild-type dPER Overexpression Does Not Cause Abnormal Circadian Phenotypes

The elevated dPER levels in the p{*dper*(OP1)} raise the possibility that the impaired clock functions might be due to high dPER levels. To rule out this possibility, I created the p{*dper*(OX)} fly strain, which carries extra copies of the p{*dper*(WT)} transgenes in addition to the endogenous *dper* gene in the *w1118* background. As shown in Figure 2.6A, the increase of *dper* copy number in the p{*dper*(OX)} flies resulted in high level of dPER that was comparable to that of the p{*dper*(OP1)} strain (Figures 2.3A-B). Locomotor activity rhythm assays showed that the p{*dper*(OX)} flies had no apparent defect in circadian behavior (Figure 2.6B). Furthermore, in contrast to the p{*dper*(OP1)} flies, mRNA levels of *dtim*, *dcwo*, and *dgol* were all significantly decreased in the p{*dper*(OX)} flies (Figure 2.6C), which is consistent with the increased repressor function of dPER due to high expression levels. This strongly suggests that the impaired clock function in the p{*dper*(OP1)} flies was caused by impaired dPER activity as a result of codon optimization rather than high dPER expression.

Mechanisms of Codon Influences on Circadian System

The impaired PER function in the circadian clock of the p{*dper*(OP1)} flies despite having identical amino acid sequence to the wild-type protein suggests that protein structure of dPER is altered. To test this possibility, I performed limited trypsin digestion assay, in which differential sensitivities can indicate protein structural changes. Fly head extracts from p{*dper*(WT)}, p{*dper*(OP1)}, and p{*dper*(OX)} flies were obtained and subjected to treatment with the same concentration of trypsin. As shown in Figures 2.7A, dPER was significantly more resistant to trypsin in the p{*dper*(OP1)} strain than the p{*dper*(WT)} strain. Importantly, dPER

trypsin sensitivities in the p{*dper*(WT)} and p{*dper*(OX)} samples were almost identical, indicating that dPER structural changes in the p{*dper*(OP1)} strain were not due to overexpression.

To further confirm our conclusion, I carried out thermal shift assay (Molina et al. 2013; Jafari et al. 2014) for the p{*dper*(WT)} and p{*dper*(OP1)} extracts. This assay quantifies changes in thermal denaturation and aggregation temperature of a protein as a result of treatment by increasing temperatures and such changes indicate structural changes of a protein. As shown in Figures 2.7B, although increasing temperatures resulted in gradual precipitation of dPER in both extracts, the precipitation rates were much higher in the p{*dper*(OP1)} extracts than those of the p{*dper*(WT)} at every temperature above 39°C (Figure 2.7b, bottom panel). Together, these results demonstrate that despite having the same amino acid sequence, dPER in the p{*dper*(WT)} and p{*dper*(OP1)} extracts are structurally different. Thus, non-optimal codons of *dper* play an important role in affecting dPER protein structures, most likely through their effects on co-translational protein folding.

Side-by-side comparison of dPER phosphorylation profiles at different time points in LD and DD indicates that the dPER is hypophosphorylated in the p{*dper*(OP1)} strain compared to that in the p{*dper*(WT)} strain (Figure 2.8A). To confirm the differences of migration were due to phosphorylation alternation, I treated the protein extracts with phosphatases, which removed all the phosphor-groups from proteins. After treatment, WT and OP1 proteins were shifted to the same position in the gel (Figure 2.8B), demonstrating the impaired phosphorylation programs in OP1 flies. The difference was more prominent for the time points during subjective night (CT

16-24), when dPER is mostly hyperphosphorylated in the p{*dper*(WT)} strain. These results further highlight the structural differences of dPER proteins in these two strains.

dPER phosphorylation has several important functional impacts, affecting its nuclear translocation, transcriptional repressor activity, and protein stability (Cyran et al. 2005; Blau 2008; Chiu et al. 2008; Kivimae et al. 2008; Ko et al. 2010; Garbe et al. 2013). The *dper* region that is optimized in the p{*dper*(OP1)} strain encodes the N-terminal part of dPER that contains major DBT phosphorylation sites that are necessary and sufficient to mediate its ubiquitination and degradation (Chiu et al. 2008). The localized codon usage effect on translation and folding raised the possibility that codon optimization in this region should result in impaired dPER phosphorylation by DBT. To test this, I compared the ability of dPER (wild-type or OP1) to be phosphorylated by DBT in cultured *Drosophila Schneider* (S2) cells. In this assay, the expression of recombinant dPER and DBT can recapitulate the DBT-dependent progressive phosphorylation and subsequent SLIMB-mediated degradation of dPER (Ko et al. 2002; Chiu et al. 2008). I first performed the assay in the presence of the proteasome inhibitor MG132 to block dPER degradation, so that the phosphorylation and degradation processes could be separated. Similar to previous results, DBT induction resulted in progressive phosphorylation of the wild-type dPER, which became mostly hyperphosphorylated after 24 hours (Figure 2.9 A). As predicted, the progressive phosphorylation process of OP1 dPER induced by DBT is markedly delayed, as indicated by the almost absence of dPER phosphorylation at 6 hrs after DBT induction, and the presence of hypophosphorylated species at later time points (Figures 2.9A-B).

When such assays were performed in the absence of MG132 (Figure 2.9C), the DBT-triggered dPER degradation was significantly impaired for dPER(OP1), indicating that dPER(OP1)

is more stable than the wild-type dPER due to impaired DBT phosphorylation (Figures 2.9C-D). This result also provides an explanation for the elevated dPER protein levels in the p{*dper*(OP1)} flies.

Phosphorylation of Serine 47 (S47) is a critical DBT phosphorylation event in the N-terminal end of dPER that generates an atypical SLIMB binding site (Chiu et al. 2008). To confirm the impact of codon optimization on DBT mediated phosphorylation *in vivo*, I compared pS47 levels at different LD time points in the head extracts of p{*dper*(WT)} and p{*dper*(OP1)} flies by using a S47 phospho-specific antibody. As shown in Figure 2.10A, the levels of S47 were markedly lower in the p{*dper*(OP1)} despite its higher dPER levels than the p{*dper*(WT)}. Together, these results demonstrate that codon optimization of N-terminal part of *dper* gene lead to altered dPER protein structure, resulting in impaired DBT phosphorylation at the N-terminal end of the dPER and less efficient protein degradation.

I further test whether this ineffective phosphorylation of p{*dper*(OP1)} is linked to altered DBT binding (Kim et al. 2007) using GST pull-down assays (Preuss et al. 2004). In this assay, DBT was purified using GST-tag from a stable *Drosophila* cell line. The amount of dPER bound to DBT was then detected by western blot analysis. The enhanced binding of p{*dper*(WT)} to DBT at ZT24 than at ZT16 (Figure 2.10B) was consistent with the progressive phosphorylation and accelerated degradation of dPER during the late night/early day (Muskus et al. 2007). Binding of p{*dper*(OP1)} to DBT was highly inefficient at ZT24, even though it is relatively abundant.

To further establish the role of codon usage in determining dPER structure, I created transgenic flies harboring p{*dper*(OP2)}, in which the *dper* region that encodes for the central

part of dPER (downstream of the PAS domains) was codon optimized (Figure 2.11A). This part of *dper* also has several regions with relative low codon usage scores and encodes for protein domains that were mostly predicted to be intrinsically disordered (Figure 2.1A). This region includes the *per*-short domain and part of the CLK/CYC inhibition domain (CCID) (Yu et al. 1987; Chang and Reppert 2003). It also contains a number of phosphorylation sites that are critical for controlling clock speed, such as T583, S585, S589, and S596 (Chiu et al. 2011). Locomotor activity assays showed that the p{*dper*(OP2)} flies also exhibited impaired behavioral rhythms as observed in p{*dper*(OP1)} flies, including reduced morning anticipation in LD and dampened rhythms in DD (Figures 2.11B-C).

Interestingly, unlike the p{*dper*(OP1)} flies, dPER protein levels were comparable between the p{*dper*(WT)} and p{*dper*(OP2)} flies in LD and DD with a similar phase (Figure 2.11D). However, dPER phosphorylation rhythms were severely impaired in the p{*dper*(OP2)} flies and dPER accumulates in hypophosphorylated forms. Furthermore, trypsin sensitivity assays showed that dPER in p{*dper*(OP2)} head extracts was significantly more resistant to trypsin digestion compared to that in the p{*dper*(WT)} (Figure 2.11E). Together, these results further demonstrate a role for codon usage in affecting dPER protein structure and function.

Summary

The *in vivo* role of codon usage was previously unclear in animal systems. In this study, I demonstrate that *dper* codon usage affects dPER structure and is critical for its function in the *Drosophila* circadian clock. Together with our previous studies in *Neurospora*, our results here

suggest that codon usage is a universal mechanism that regulates protein structure and function from fungi to animals.

The role of codon usage in regulating dPER protein structure and function is supported by several lines of evidence. First, codon optimization of part of *dper* open reading frame resulted in severe impairment of circadian locomotor activity rhythms and dPER function in the circadian negative feedback loop. Second, *dper* codon optimization affect the ability of dPER to interact with CLK/CYC, providing a mechanism for the impaired circadian negative feedback loop in the optimized flies. Third, despite having the same amino acid sequence, dPER protein in the codon optimized flies exhibited significantly differential sensitivities to partial trypsin digestion and in thermal shift assays compared to that extracted from flies with the wild-type gene, indicating dPER structure changes due to codon manipulation. Furthermore, *dper* codon optimization impaired dPER phosphorylation at the site of codon changes, which led to altered dPER stability. Finally, I showed that the observed effects of *dper* codon optimization is not due to dPER protein overexpression. Together, these results, to our knowledge, establish the first *in vivo* example in an animal system that demonstrates the role of codon usage in determining protein structure and function.

How does codon usage influence dPER structure and function? By comparing mRNA translation elongation speed in *Neurospora*, our lab demonstrated that codon usage regulates speed of elongation: preferred codons speed up translation elongation while unpreferred codons slow it down (Yu et al. 2015). Although the effects of codon usage on elongation rate have not been demonstrated in insects, a similar mechanism should exist due to the conservation of the translation process. Thus, codon usage affects the amount of time available for co-translational

folding. This is consistent with previous studies that translation rate and synonymous codon usage can affect protein folding and functions (Komar et al. 1999; Zhang et al. 2009; Siller et al. 2010; Spencer et al. 2012; Kim et al. 2015; Presnyak et al. 2015). I have previously observed genome-wide correlations that optimal codons are preferentially used in regions that are predicted to be well folded while relatively more unpreferred codons are used in protein regions predicted to be unstructured. Our studies and previous uncovered correlations between codon usage and certain protein structural motifs suggest that codon usage and protein structures co-evolve and are adapted to each other (Zhou et al. 2009; Pechmann and Frydman 2013; Pechmann et al. 2014; Zhou et al. 2015). Therefore, there is a codon usage code within genetic codons that generates elongation speed rhythms to optimize the cotranslational folding process.

The predicted unstructured or intrinsically disordered protein (IDP) domains widely exist in all analyzed proteomes. Despite the fact that they are not predicted to form stable three-dimensional structures, IDPs have been shown to play important roles in many biological processes (Dyson and Wright 2005; Dunker et al. 2008; Tompa 2011). Similar to *Neurospora* FRQ, most of the *Drosophila* PER protein is predicted to be unstructured. Here I showed that codon optimization of parts of *dper* that encode for N-terminal and middle part of the unstructured regions resulted in altered dPER structure and function. Therefore, both our previous studies in *Neurospora* and this current study highlight the importance of codon usage in the unstructured protein regions. These putative IDP regions may require relatively longer cotranslational folding time either to fold into certain structures or to serve as platforms for inter- or intramolecular protein-protein interactions.

Post-mitotic cells, especially neurons, seem to be extremely sensitive to misfolding proteins. In fact, neural tissues are under the almost strongest mistranslation-induced protein misfolding constraints on their coding sequence evolution (Drummond and Wilke 2008). It has been shown that disruption of translational fidelity can impair the normal functions of terminally differentiated neurons. Actually, several neurodegenerative diseases are caused by the misfolded proteins produced by defective translation machinery (Iwatsubo 2000; Ross and Poirier 2004; Lee et al. 2006). For example, either low levels of mischarged tRNAs (Lee et al. 2006), or mutations in Aminoacyl-tRNA synthetases (Antonellis and Green 2008), can lead to an intracellular accumulation of misfolded proteins in neurons, eventually leading to neurodegenerative diseases. The basis of behavioral circadian rhythms is a brain neuronal circuit (Nagoshi et al. 2010). In our study, changes of synonymous codons in *dper* sequence results in abnormal circadian behavioral outputs, suggesting a circadian neuron dysfunction. Therefore, maintaining the equilibrium between mRNA codons and tRNA anticodons (Schmitt et al. 2014) is extremely critical to produce functional gene products in neural systems. This is also consistent with a previous bioinformatics analysis that tRNA pools are adapted to gene expression in a tissue-specific manner (Waldman et al. 2010), and this correlation in whole brain is ranked among top when compared to various tissues.

Synonymous codon mutations have been associated with many human diseases with unknown mechanisms. These diseases include cystic fibrosis, amyotrophic lateral sclerosis (ALS) and Chron's disease (Bartoszewski et al. 2010; Lazrak et al. 2013b; Bali and Bebok 2015; Liu et al. 2016). A single synonymous SNP in the form of a rare codon in the human *multidrug resistance 1 (MDR1)* gene was found to result in altered drug and inhibitor interactions (Kimchi-

Sarfaty et al. 2007). For circadian clock related diseases, a T2434C synonymous polymorphism in exon 18 of hPER1 was found to be associated with extreme diurnal performance (Carpen et al. 2006) and a G2114A in hPER2 were linked to diurnal preference (Matsuo et al. 2007). Our results here suggest that the change of codon usage in these genes may be a mechanism that contributes to impaired function of the encoded proteins.

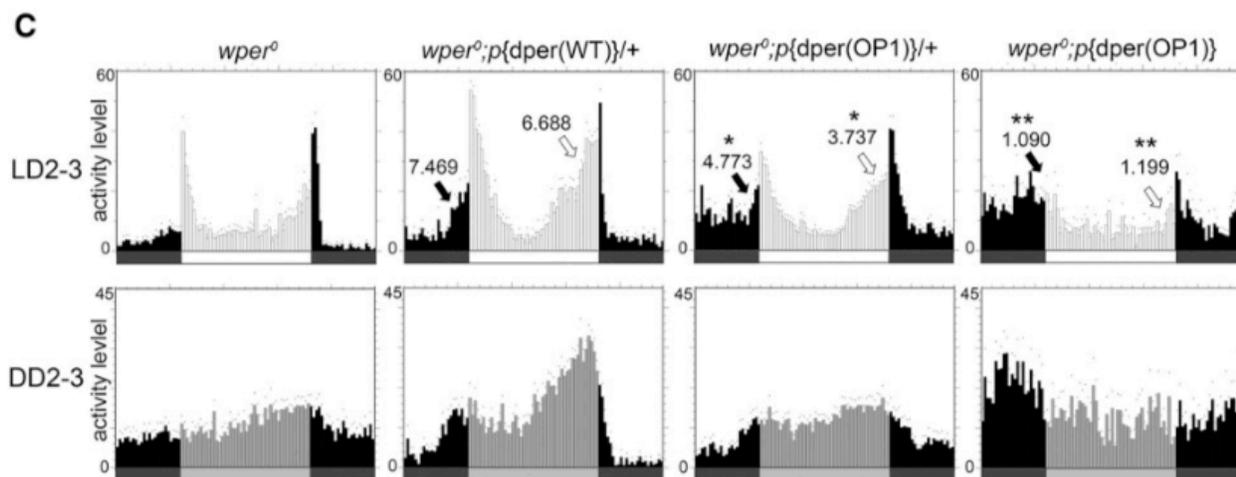
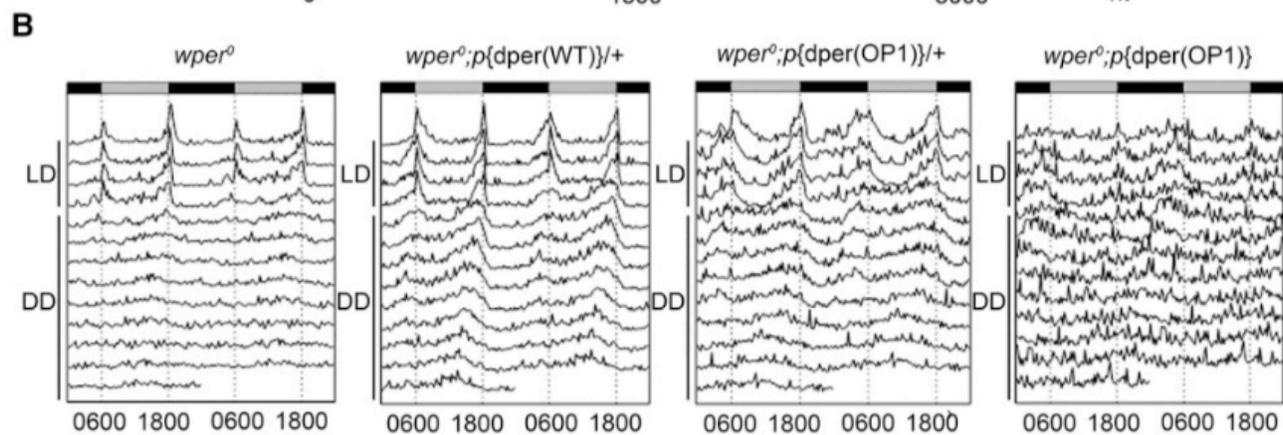
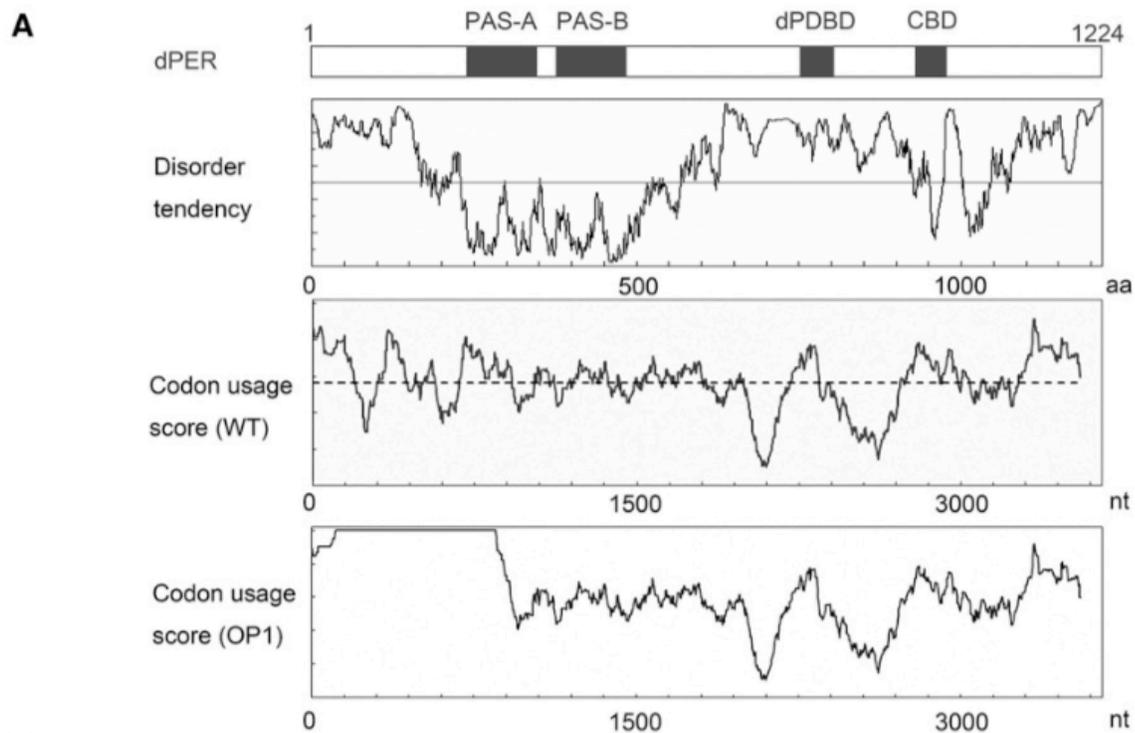


Figure 2.1 Codon optimization of the N-terminal part of dper led to impaired circadian locomotor activity rhythms.

A) from top to bottom) A diagram depicting the previously identified domains of dPER. PAS-A and PAS-B (PAS domains); (dPDBD) DBT-binding domain; (CBD) dCLK-binding domain. Disorder tendency plot of the dPER protein using IUPred. Codon usage score plot (CAI value, window 35) of wild-type dper. Codon usage score plot (CAI value, window 35) of dper(OP1). The dashed line in the codon usage for the wild-type gene indicates the average CAI of wild-type dper.

B) Double plot actograms showing locomotor activity rhythms of the $wper^0; p\{dper(WT)\}$ and $wper^0; p\{dper(OP1)\}$ fly strains in 4 d of light/dark cycles (LD) and 7 d of constant darkness (DD). C) Education graphs generated from locomotor activity analysis showing the rhythms of the indicated strains. The Y-axis represents activity levels. (Top) The activity data generated by averaging the second and third days in light/dark cycles (LD 2–3). (Bottom) The activity data generated by averaging the second and third days in DD (DD 2–3). Arrows indicate morning anticipation (black) and evening anticipation (white) behaviors with their respective anticipation index (AI) values. The statistical analysis was performed using a two-tailed t-test to compare the AIs between the OP1 mutants and the wild-type per gene rescue strain. (*) P-value < 0.05; (**) P-value < 0.01.

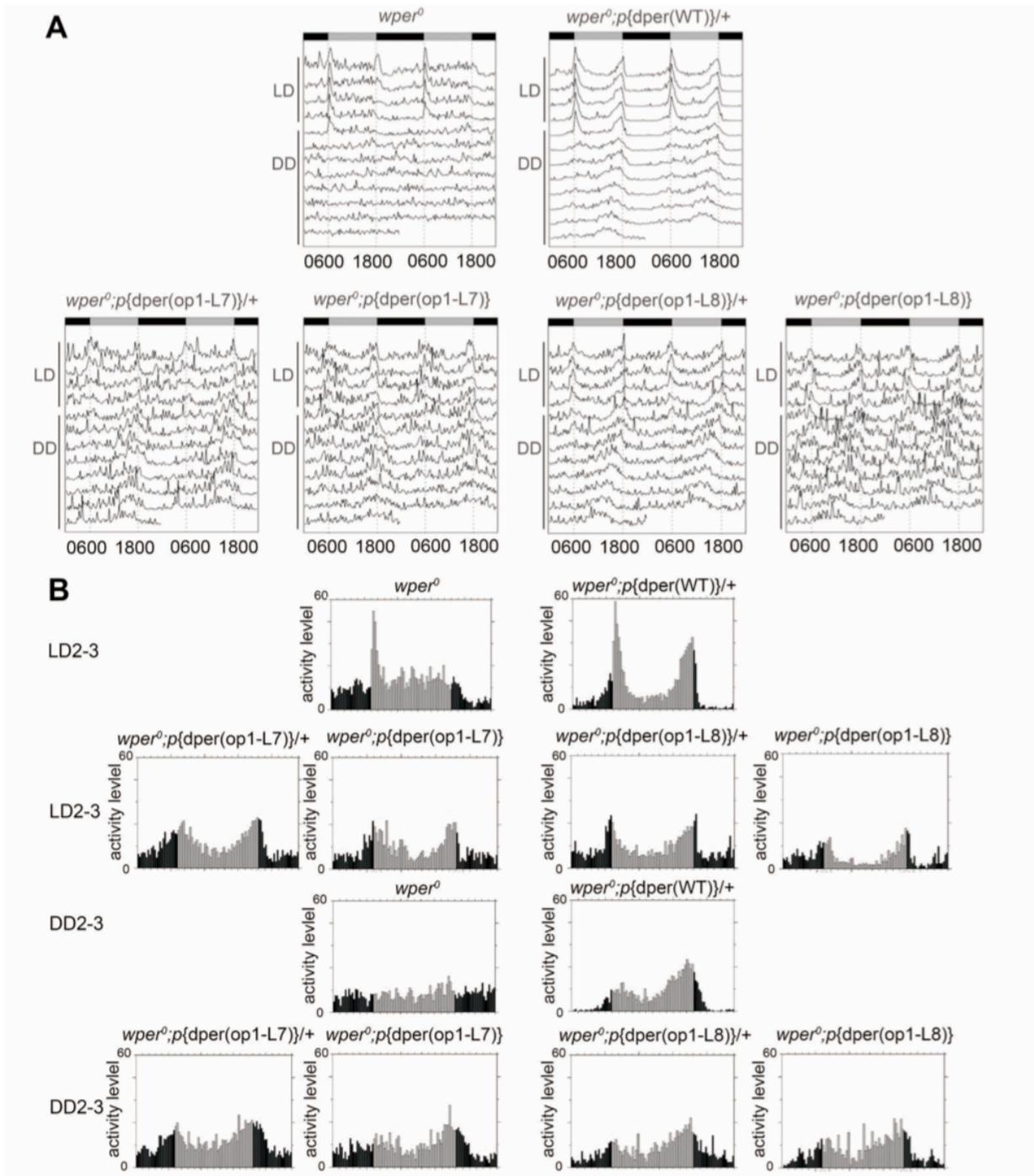


Figure 2.2 Impaired locomotor activity rhythms of additional OP1 strains.

A) Doubleplot actograms showing locomotor activity rhythms of the $wper0: p\{dper(WT)\}$, $wper0: p\{dper(OP1-L7)\}$, and $wper0: p\{dper(OP1-L8)\}$ fly strains in 4 d of LD and 7 d OF dd.

B) Education graphs generated from locomotor activity analysis showing the rhythms of the indicated strains. Top: the activity data generated by averaging the second the third days in light/dark cycles (LD 2-3). Bottom: the activity data generated by averaging the second and third days in DD (dd 2-3).

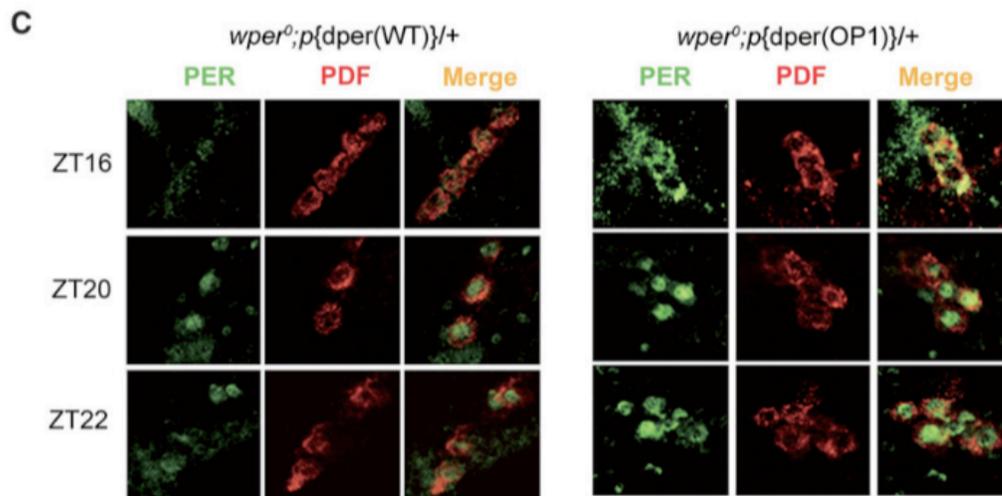
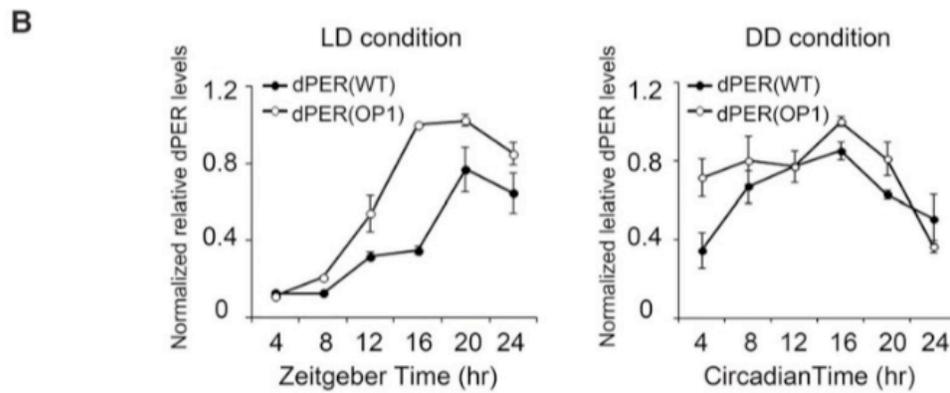
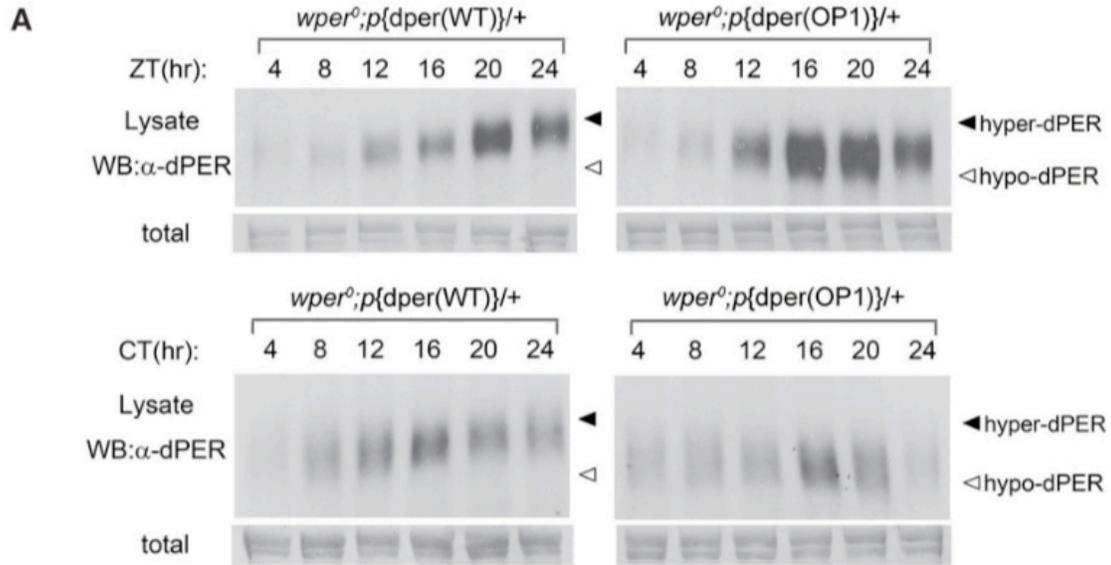


Figure 2.3 Impaired dPER rhythms in OP1 flies.

A) Western blot results showing the dPER molecular rhythm in LD (top) and DD (bottom) for wild-type and OP1 flies. The filled and open arrowheads indicate the hyperphosphorylated and hypophosphorylated dPER proteins, respectively. Membrane staining was used as a loading control.

B) Densitometric analyses of the results from three independent experiments. The levels of dPER were normalized to the loading control. Error bars indicate \pm SD.

C) Immunohistochemistry assay of dPER expression in pigment dispersing factor (PDF)-positive (PDF⁺) circadian neurons in fly brains. Adult flies were entrained to LD cycle, and brains were dissected for immunohistochemistry analysis at the indicated time points.

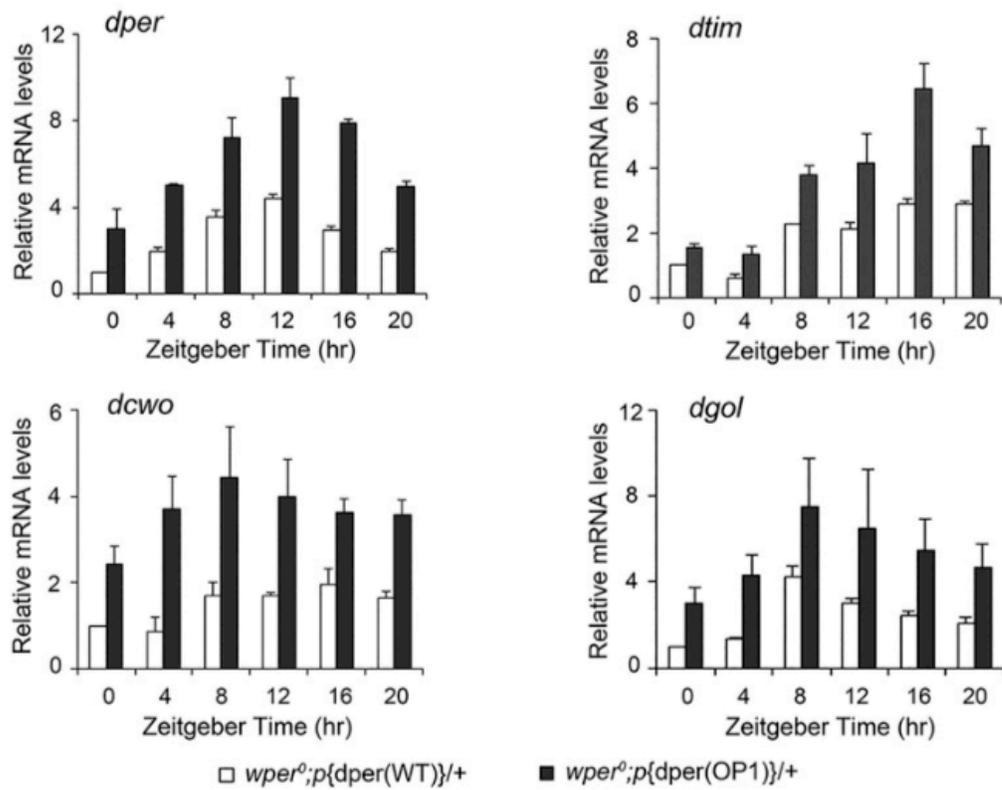


Figure 2.4: Impaired circadian negative feedback loop in OP1 flies.

Quantitative RT-PCR assays showing the mRNA levels of *dper*, *dtim*, *dcwo*, and *dgol*. Error bars indicate SD.

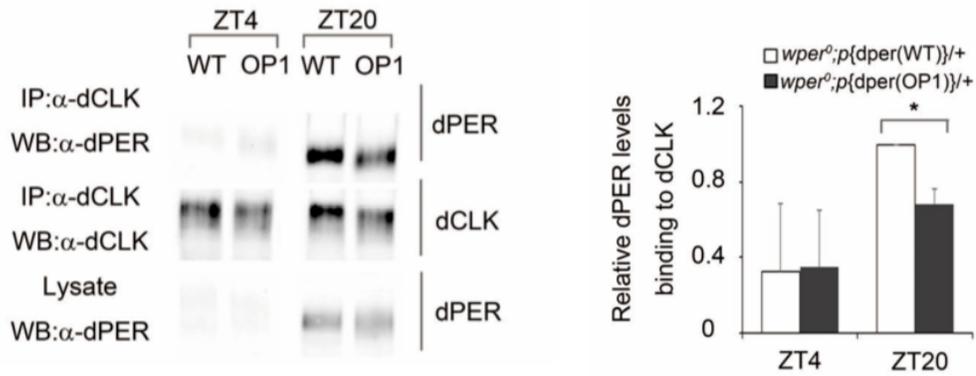
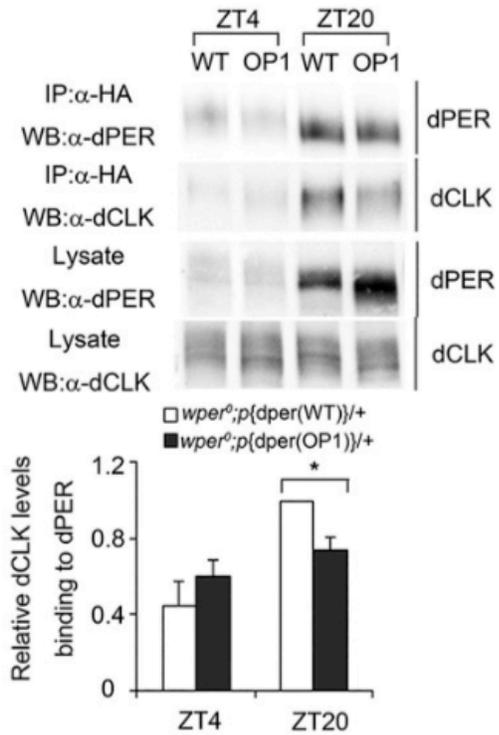
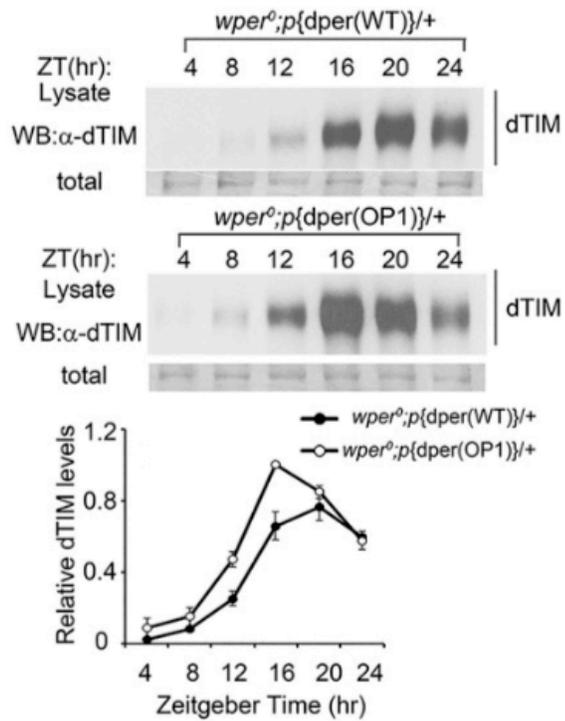
A**B****C**

Figure 2.5: PER-CLK interaction is impaired in $wper^0; p\{dper(OP1)\}$ flies.

A) Left: Representative results of co-immunoprecipitation assay using CLK antibody showing the reduced interaction between dPER and CLK in $wper^0; p\{dper(OP1)\}$. Head extracts were prepared from $wper^0; p\{dper(WT)\}$ and $wper^0; p\{dper(OP1)\}$ flies collected at the indicated times (ZT). Aliquots containing equal amount of total protein for each sample were used for immunoprecipitation with anti-CLK (Santa Cruz Biotechnology). Right: Bar graph displaying the amount of dPER interacting with CLK.

B) Immunoprecipitation assay showing the reduced interaction between dPER and CLK in the $wper^0; p\{dper(OP1)\}$ flies. Head extracts were prepared from $wper^0; p\{dper(WT)\}$ and $wper^0; p\{dper(OP1)\}$ flies collected at the indicated times (ZT). (Top) Representative Western blot results are shown. (Bottom) Densitometric analyses from four independent biological experiments. The amount of dCLK was normalized to the HA (dPER) signal in the immunoprecipitation.

C) Western blot analysis showing the protein levels of TIM in the indicated fly strains in LD. Membrane staining was used as a loading control. (Bottom) Densitometric analyses of the Western blot results. Error bars indicate \pm SD.

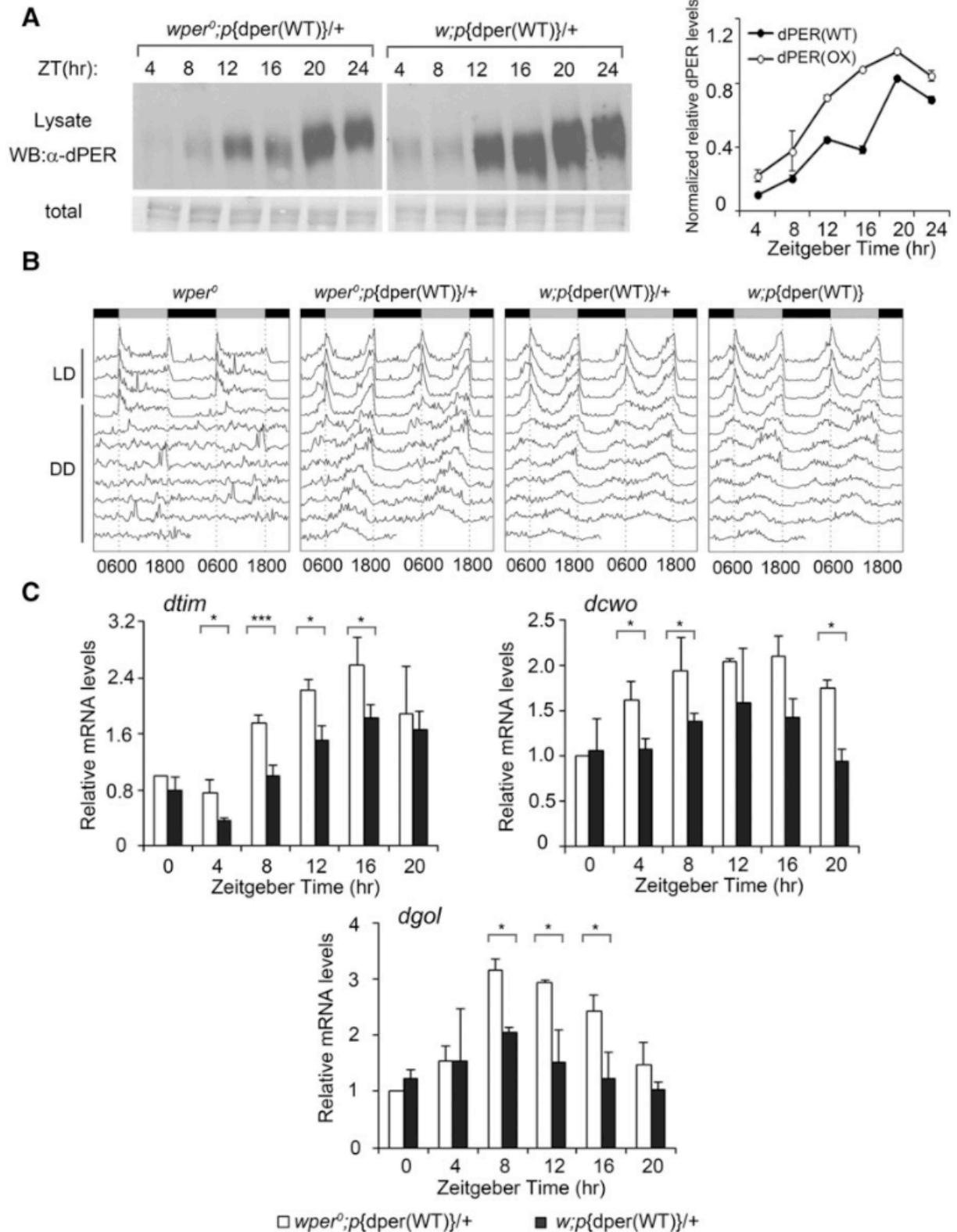


Figure 2.6: Overexpression of wild-type dPER does not result in phenotypes that resemble OP1 flies.

A) left panels: Western blot analysis shows that the levels of dPER were elevated to levels comparable with those of OP1 strains in the $w;p\{dper\}$ (WT) (OX) fly strains due to the extra copy number of wild-type dPER. Note that endogenous per is located on the X chromosome. Membrane staining was used as a loading control. Right panels: Densitometric analyses of the Western blot results. Error bars indicate \pm SD.

B) Double plot actogram showing circadian locomotor activity rhythms of the indicated strains in 4 d of LD and 7 d of DD.

C) Quantitative RT-PCR assays showing the mRNA levels of *dtim*, *dcwo*, and *dgol* in the indicated strains. Error bars indicate \pm SD. (*) $P < 0.05$

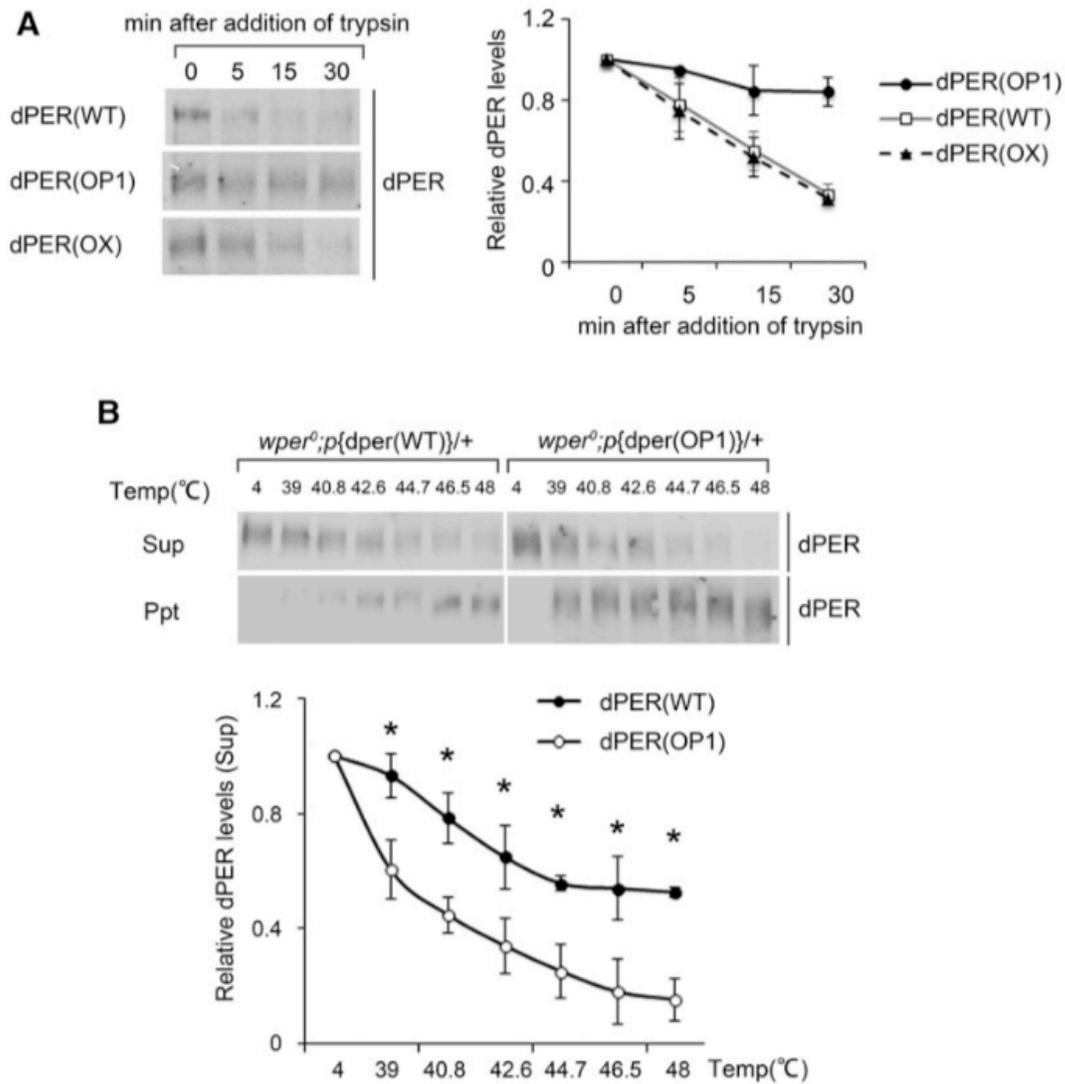


Figure 2.7: Codon optimization of *dper* results in altered dPER sensitivity to trypsin digestion and heat treatment.

A) left panels: Western blots showing the levels of dPER from the indicated strains after partial trypsin (0.5 $\mu\text{g}/\text{mL}$) digestion at the indicated time points. Right panels: Densitometric analyses of the Western blot results from three independent experiments. The levels of dPER at time point 0 were set as 1.

B) Thermal shift assays comparing the sensitivity of dPER from the indicated strains to heat treatment. (Top panels) Western blots showing the levels of dPER in the supernatant (top blot) or precipitate (bottom blot) from wild-type and OP1 strains. Bottom panels: Densitometric analyses of the results from three independent experiments. The levels of dPER at 4°C were set as 1. Error bars indicate $\pm\text{SD}$. (*) $P < 0.05$.

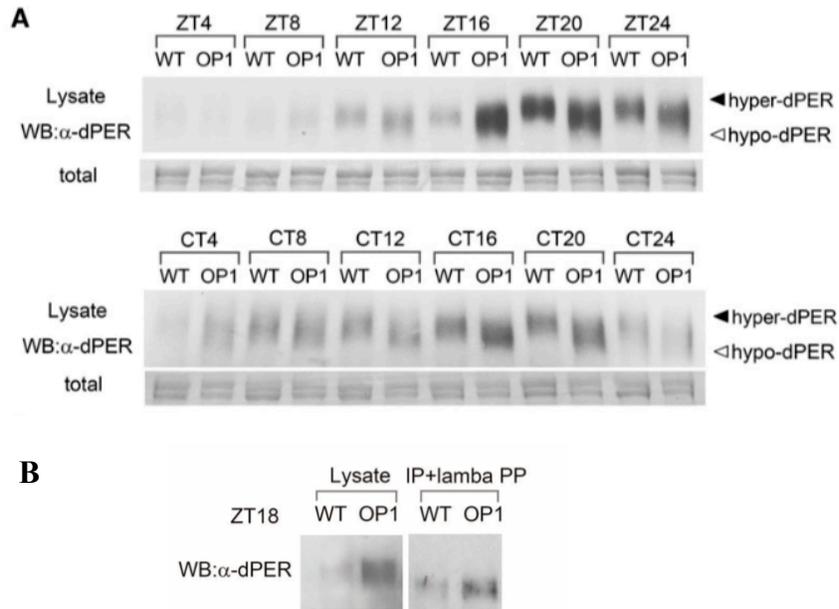


Figure 2.8: Impaired dPER phosphorylation profiles and degradation in OP1 flies.

A) Western blots showing a side-by-side comparison of dPER phosphorylation profiles at different time points in LD and DD between the $wper^0;p\{dper(WT)\}$ and $wper^0;p\{dper(OP1)\}$ flies. Membrane staining was used as a loading control.

B) dPER immunoprecipitated from fly head samples at ZT18 was subjected to in vitro phosphatase treatment, followed by western blot assay. Treatment with lambda phosphatase made optimized dPER indistinguishable from wild-type dPER, confirming that differences in electrophoretic migration was caused by different phosphorylation program in $wper^0;p\{dper(WT)\}$ and $wper^0;p\{dper(OP1)\}$ flies.

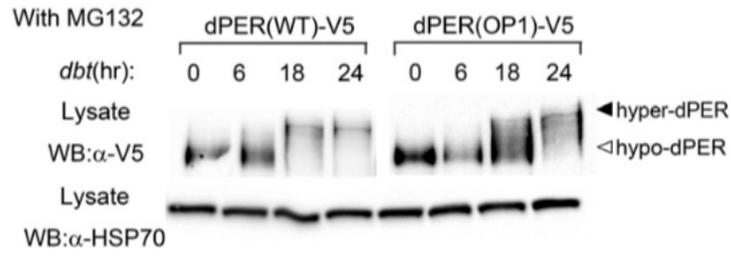
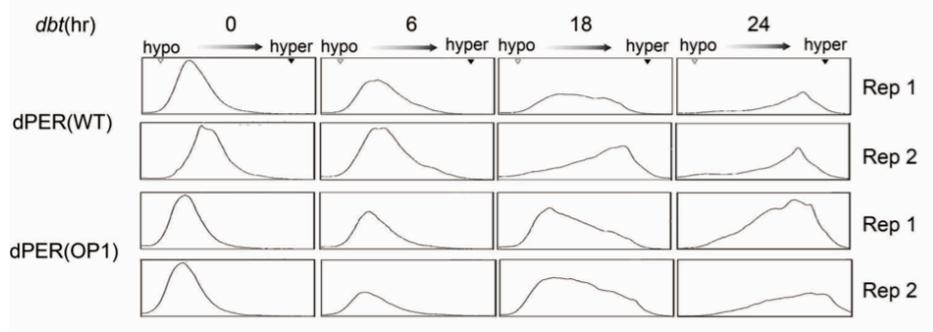
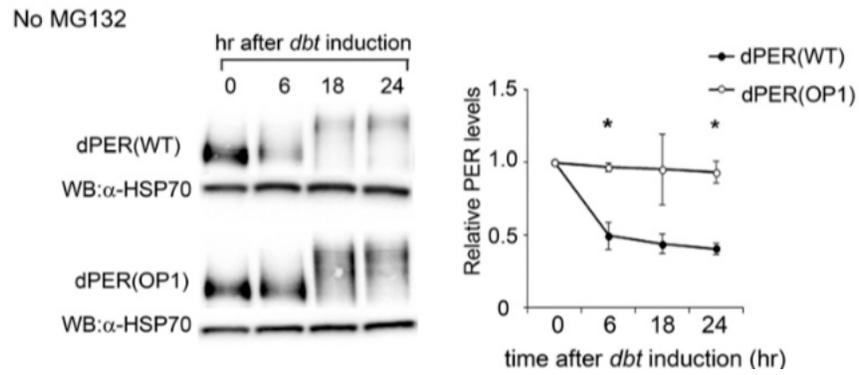
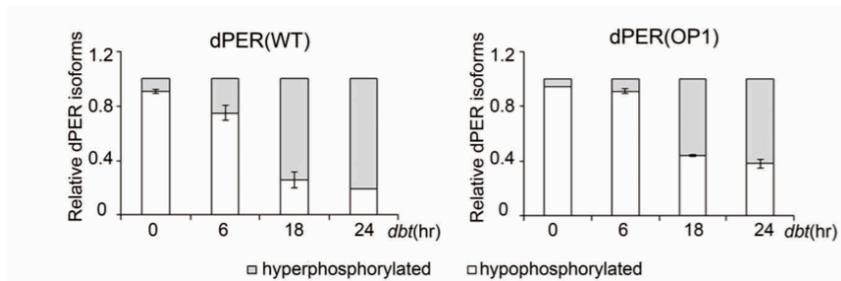
A**B****C****D**

Figure 2.9: dPER phosphorylation profiles and degradation in OP1 flies.

A) *Drosophila Schneider* (S2) cells were cotransfected with dbt and dper (pAC-dper-V5) variants and collected at the indicated times (hours) after dbt induction, the culture medium contained MG132 to inhibit dPER degradation. HSP70 signal was used as a loading control.

B) Qualification of different phosphorylation isoforms from in vitro phosphorylation results in A). The peak of signals was shifted to the right as phosphorylation progressed. OP1 dPER phosphorylation by DBT was markedly delayed, as indicated by the slower shift of peak to the right when compared to the wild-type dPER.

C) *Drosophila Schneider* (S2) cells were cotransfected with dbt and dper (pAC-dper-V5) variants and collected at the indicated times (hours) after dbt induction without MG132. HSP70 signal was used as a loading control.

D) Qualification of the relative ratio of hypophosphorylated and hyperphosphorylated part of dPER against total dPER protein signals at each timepoint shown in C). More wild-type dPER was presented as hyperphosphorylated isoforms. Densitometric analyses of the results were from three independent experiments.

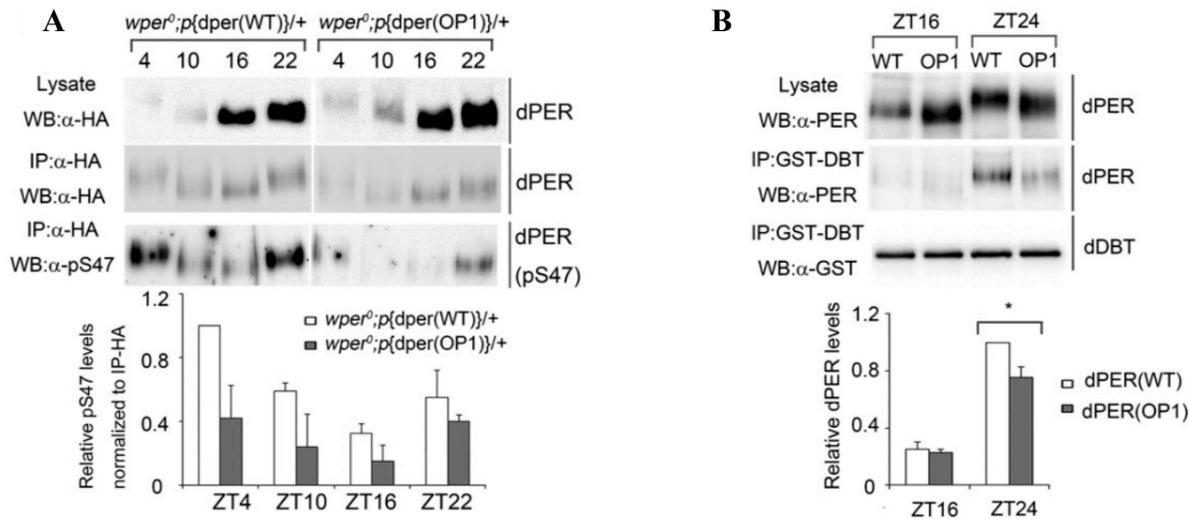


Figure 2.10: Codon optimization changes dPER local structure and protein-protein interaction, which causes altered phosphorylation profiles of OP1 dPER.

A) Western blot analysis using anti-pS47 antibody showing the reduction of S47 phosphorylation of dPER in the *wper⁰;p{dper(OP1)}* flies. Head extracts were prepared at the indicated times (ZT). dPER-HA-containing immune complexes were recovered using anti-HA beads, and dPER(S47) were detected by Western blots using an anti-pS47 antibody. (Bottom) Densitometric analyses of the results from three independent experiments. Error bars indicate \pm SD.

B) GST pull-down assay showing the reduced interaction between dPER and DBT in the *wper⁰;p{dper(OP1)}* flies. Head extracts were prepared from *wper⁰;p{dper(WT)}* and *wper⁰;p{dper(OP1)}* flies collected at the indicated times (ZT). (Top) Representative Western blot results are shown. (Bottom) Densitometric analyses from four independent biological experiments. (*) $P < 0.05$.

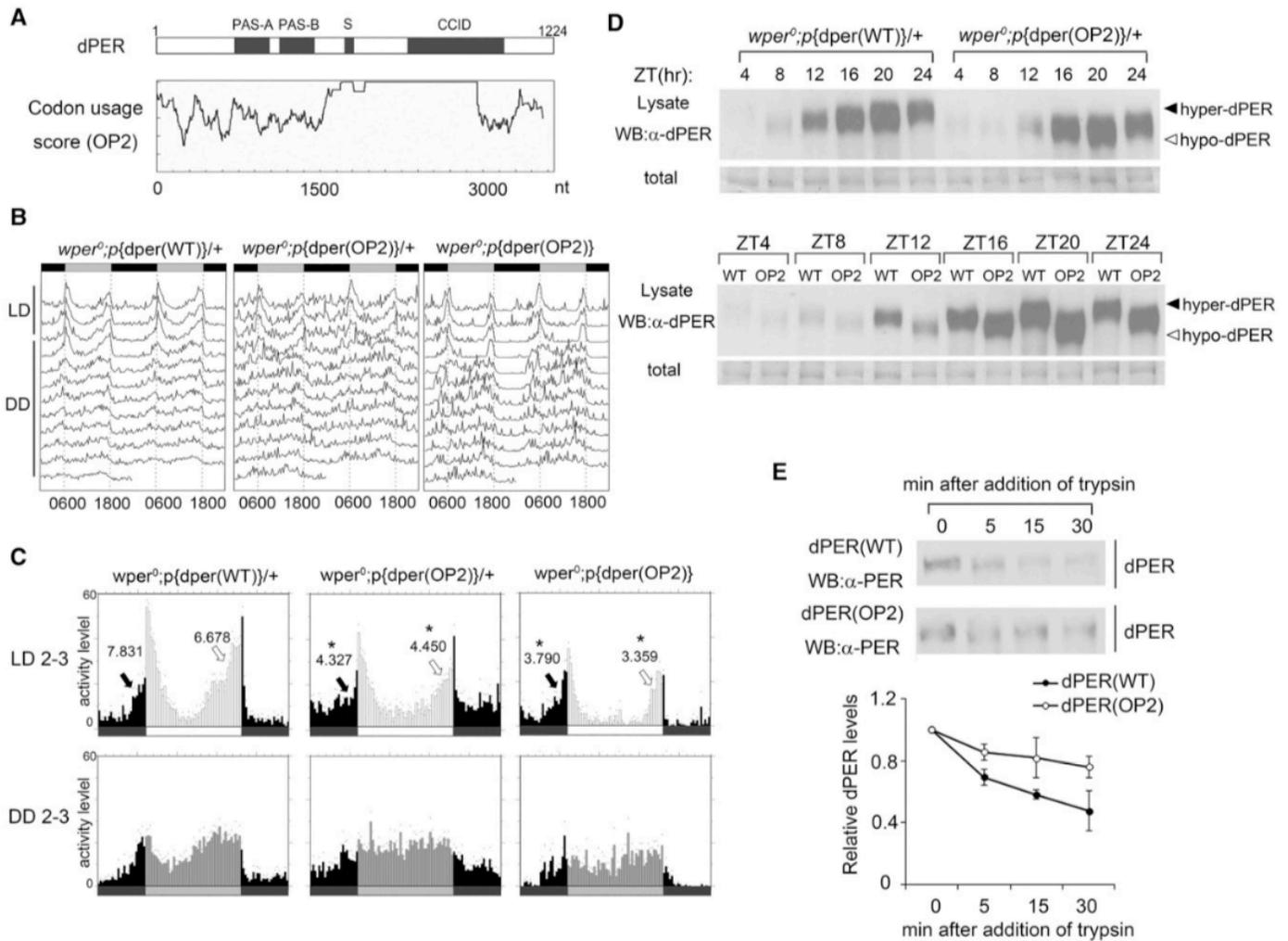


Figure 2.11: Codon optimization of the central part of dper resulted in impaired circadian rhythms and altered dPER structure.

A) Diagrams showing the dPER protein domains and the codon usage score plot of dper (CAI value, window 35) after codon optimization.

B) Double plot actogram showing the circadian rhythms of $wper^0;p\{dper(WT)\}/+$ and $wper^0;p\{dper(OP2)\}/+$ strains after 4 d of LD and 7 d of DD.

C) Education graphs generated from locomotor activity analysis showing the circadian rhythms of the indicated strains in LD 2-3 (top) and in DD 2-3 (bottom). Arrows indicate morning anticipation (black) and evening anticipation (white) behaviors with their respective AI values.

(*) $P < 0.05$.

D) top panels: Western blot results using dPER antibody showing the dPER rhythm in LD.

(Bottom panels) Side-by-side Western blot analysis results showing the dPER mobility

differences between two fly strains. Membrane staining was used as a loading control. (E, top

panels) Western blots comparing the sensitivity of dPER from the indicated strains with partial

trypsin (0.5 μ g/mL) digestion. Bottom panels: Densitometric analyses of the Western blot results

from three independent experiments. The levels of dPER at time point 0 were set as 1. Error bars

indicate \pm SD.

Circadian parameters

	Num Total	Num Alive	%Rhythmic	Tau	Power
wper ⁰	32	32	0	***	***
wper ⁰ ;p {dper(WT)}/+	32	32	93.8	23.8	100
wper ⁰ ;p {dper(OP1)}/+	16	15	80	23.4	89.4
wper ⁰ ;p {dper(OP1)}	18	11	27.3	22.7	51.2
wper ⁰ ;p {dper(OP1-L7)}/+	26	25	76	23.8	83.9
wper ⁰ ;p {dper(OP1-L7)}	32	30	56.7	23.8	77.3
wper ⁰ ;p {dper(OP1-L8)}/+	20	19	78.9	23.2	79.3
wper ⁰ ;p {dper(OP1-L8)}	14	14	57.1	23.2	70.5
wper ⁰ ;p {dper(OP2-L1)}/+	32	32	87.5	23.6	89.2
wper ⁰ ;p {dper(OP2-L1)}	32	31	68.2	23.1	78.3
wper ⁰ ;p {dper(OP2-L4)}/+	16	15	80	23	87.8
wper ⁰ ;p {dper(OP2-L7)}/+	32	31	83.9	22.7	87
wper ⁰ ;p {dper(OP2-L7)}	32	28	75	22.3	75.1
w;p {dper(WT)}/+	32	27	93.6	22.8	94.5
w;p{dper(WT)}	32	25	76	23.2	93.3

Table 2.1 Circadian behavioral parameters of the indicated fly strains. Additional independent experiments were performed. Representative results were shown.

CHAPTER THREE

CODON OPTIMALITY REGULATES RAS EXPRESSION AND STRUCTURE

Codon Usage Strongly Affects Expression of KRAS

The average codon adaptation index (CAI) of *HRAS* is 0.88, whereas *KRAS*, which uses many rare codons, has an average CAI of 0.71. (Figure 3.1A and Lampson et al, 2013). For example, GTG and ATC are the most preferred codons for valine and isoleucine, respectively, in the human genome. Although these optimal codons are overwhelmingly used in *HRAS*, they are rarely used in *KRAS* (Figure 3.1B). Similar to what previously reported (Lampson et al. 2013), transient transfection of an N-terminally FLAG-tagged wild-type *HRAS* or *KRAS* cDNA (referred to as Hras and Kras constructs, respectively) in human embryonic kidney HEK-293T cells resulted in highly divergent protein expression (Figure 3.1C). However, as noted above, codon bias can affect the entire process of protein production, and hence I proposed to dissect the contribution of each to the ultimate end product of a functional protein. To this end, I first needed to benchmark the effect of the rare codons on KRAS expression. I thus chose three previously created versions of human *KRAS* cDNA for this analysis (Lampson et al. 2013). The wild-type *KRAS* was used as the control for native codon usage. In opKRAS, the most nonoptimal valine (GTA) was changed to the most common (GTG), whereas Isoleucine is changed from ATA to ATC. For *KRAS**, the *KRAS* cDNA was optimized to mimic *HRAS* sequence (Figure 3.2B). As expected, transient transfection of vectors encoding each of these three cDNAs into human 293T cells yielded a stepwise increase in protein levels. Codon

optimization of only 18 codons in opKRAS resulted in an approximately 10-fold increase in KRAS protein level compared to the wild-type KRAS construct (Figure 3.2B). Remarkably, Kras* produced about 100-fold more KRAS over the control (Figure 3.2B), which is similar to the difference between proteins produced from Hras and Kras cDNAs (Figure 3.1C).

I note here that changing progressively more rare codons to common in *KRAS* increases KRAS protein levels, but the effect of codon usage within specific regions of KRAS had not been explored. Thus, I further manipulated the opKras sequence, generating three partially optimized versions called N-opKras, M-opKras, and C-opKras, in which only the valine and isoleucine codons encoding the N-terminal, central, or C-terminal regions were optimized (Figure 3.3A, Table 3.1). These three versions of opKras expressed KRAS protein at levels intermediate between those of the Kras and opKras constructs (Figure 3.3B). I also mutated the opKras construct so that only nine rare isoleucine or nine rare valine codons (opKras-I and opKras-V) were optimized. Both opKras-I and opKras-V resulted in KRAS protein levels that were higher than those produced from the wild-type Kras but lower than those from opKras (Figure 3.3C, Table 3.1). Thus, the three versions of KRAS exhibited the expected effect on protein production and different synonymous codons have independent and accumulative effects on KRAS expression.

Rare Codons Suppress KRAS mRNA Translation

Comparison of the protein decay rates after the addition of protein synthesis inhibitor cycloheximide revealed that codon optimization in *opKRAS* and *KRAS** did not affect KRAS protein stability (Figure 3.4). Changes of codon should affect other aspects. Codon usage has

been shown to regulate translation elongation rate and co-translational protein folding (Zhou et al. 2013; Pechmann et al. 2014; Yu et al. 2015) and has been proposed to influence translation efficiency and accuracy (Akashi 1994; Drummond and Wilke 2008; Hershberg and Petrov 2008; Gingold and Pilpel 2011; Plotkin and Kudla 2011; Qian et al. 2012). In agreement, other lab previously reported that changing rare codons to common increased *KRAS* mRNA in the polysome fractions (Lampson et al. 2013). To determine the effect of codon usage on *KRAS* translation, I synthesized different versions of *KRAS* mRNA by *in vitro* transcription; mRNAs had 5' caps and poly(A) tails. Equal amounts of mRNA were transfected individually into 293T cells and the amount of KRAS protein produced was determined. The amount of protein produced from *opKRAS* and *KRAS** mRNAs was about 0.5- and about 4-fold higher, respectively, than that produced from the wild-type *KRAS* mRNA (Figure 3.5A). I note there that the differences in protein levels using *KRAS* mRNA were much less than those observed (transient transfection DNA vector is not the same as that used in in-vitro transcription were transfected into cells (Figure 3.2B), suggesting that additional mechanism(s) mediate the codon usage effect on protein production.

To further confirm the effect of codon usage on translation, I performed *in vitro* translation assays using 293T cell lysates. Similar fold differences in KRAS protein production were observed for the *opKRAS* and *KRAS** mRNAs (Figure 3.5B) as were observed in the above cell-based assay. Interestingly, when the same mRNAs were translated in an extract made from budding yeast, an organism with A/T-biased codons, the trend in expression pattern was opposite: The highest level of KRAS was synthesized from the wild-type *KRAS* mRNA (Figure 3.5C). Together, these results suggest that rare codons in *KRAS* suppress KRAS translation in

human cells. Furthermore, I generated stably transfected HEK-293T cell lines with Kras and with Kras* expression constructs and performed polysome profiling (Figure 3.6). Northern blot analysis of the polysomal fractions showed that the wild-type Kras mRNA was peaked in the monosome fraction (fraction #8), whereas Kras* mRNA was enriched in the polysome fractions (fractions 15-19) (Figure 1D). Again, these data indicate that optimal codons of *KRAS* promote efficient mRNA translation in cells. However, I estimate a ~4 fold increase in translation (Figures 3.5A-B), which when benchmarked against the very high amount of protein produced by *KRAS** suggest that translation is only one aspect accounting for the effect of rare codons on *KRAS* protein levels.

KRAS Codon Optimization Increases mRNA Level but does not Affect mRNA Stability

As noted above, multiple experiments support that rare codons impact *KRAS* protein production beyond translation. I had previously reported that ectopic *KRAS** generated roughly 2.5 fold more mRNA compared to wild-type *KRAS*, as assessed by qRT-PCR (Lampson et al. 2013). To more accurately measure the effect of codon usage of *KRAS* mRNA levels I utilized Northern hybridization analysis using a probe that anneals to the common 5' untranslated region of all three ectopic *KRAS* mRNAs. In cells that were transfected with different codon-optimized versions of *KRAS* cDNA expression constructs the codon optimization resulted in about 4- and 10-fold increases in mRNA levels for *opKRAS* and *KRAS**, respectively, compared to wild-type *KRAS* mRNA levels (Figure 3.7A). These results suggest that the effect of codon usage on *KRAS* mRNA expression is more pronounced than first thought and plays a major role in regulating *KRAS* expression.

Codon usage was previously shown to affect mRNA levels by influencing mRNA stability in different organisms (Presnyak et al. 2015; Bazzini et al. 2016; Boel et al. 2016; Mishima and Tomari 2016; Radhakrishnan et al. 2016; Zhao et al. 2017). To determine whether the effect of codon usage on *KRAS* mRNA levels is due to its effects on mRNA stability, I compared mRNA decay rates of wild-type and codon-optimized mRNAs after the addition of amanitin, a commonly used transcription inhibitor. Northern blot quantifications revealed that there were no significant differences in mRNA stability (Figure 3.7B). These results indicate that in addition its effect on translation efficiency, *KRAS* codon usage also has a major role in mRNA levels without overtly affecting mRNA stability. Consistent with this conclusion, it was previously shown that mammalian genes with high GC contents, which is associated with more preferred codons, have higher expression levels than those with lower GC content without affecting mRNA degradation rates (Kudla et al. 2006; Krinner et al. 2014; Newman et al. 2016).

Using the RNA-seq results from the Genotype-Tissue Expression (*GTEx*) Program, I compared the relative RNA levels of *KRAS* and *HRAS* in different human tissues. As shown in Figure 3.7C, across all tissues examined, the *KRAS* mRNA levels are much lower than those of *HRAS* mRNA. This result is consistent with the notion that regulation of mRNA level is a major mechanism that suppresses *KRAS* expression in human cells.

KRAS Codon Usage Regulates Transcription and Chromatin Modifications

The increase in *KRAS* transcript levels by codon optimization prompted us to examine *KRAS* transcription. Using human 293T cell lines stably transfected with a vector encoding wild-type *KRAS*, op*KRAS*, or *KRAS**, I performed Br-UTP-coupled nuclear run on assays to

examine transcriptional by RNA polymerase II (Pol II) at the locus of interest. Since this assay quantifies the frequency of transcription initiation, the levels of newly synthesized transcripts should not be influenced by RNA stability. Codon optimization resulted in significant increases of *KRAS* transcription with about 10-fold higher levels of transcript from *Kras** than wild-type *Kras* (Figure 3.8A).

After transcriptional initiation, the Pol II carboxy-terminal domain (CTD) is phosphorylated at serines 2 and 5 (Komarnitsky et al. 2000; Hsin and Manley 2012). Ser 5 phosphorylation of the CTD tail occurs soon after initiation, whereas Ser 2 phosphorylation of the CTD of Pol II takes place during transcriptional elongation process. To confirm the effect of codon usage on *KRAS* transcription, I compared the enrichment of phosphorylated Pol II CTD on the wild-type and codon-optimized *KRAS* by chromatin immunoprecipitation (ChIP) assays. Codon optimization resulted in a significant enrichment of both Ser 2 and Ser 5 phosphorylated Pol II at the *KRAS* transgene loci (Figure 3.8B). These results indicate that codon usage impacts *KRAS* transcription.

To determine the mechanism by which codon usage affect *KRAS* transcription, I first performed histone H3 ChIP. The occupancies of histone H3 at the *KRAS* transgene loci showed no significant differences among three stable cell lines (Figure 3.8C), suggesting that codon usage does not influence nucleosome density. I then performed ChIP assays for several histone modification marks associated with transcriptionally active chromatin. Both H3K4 trimethylation and H3K9 acetylation were enriched at the *opKRAS* and *KRAS** compared to the wild-type *KRAS* transgene locus (Figure 3.8D), consistent with mRNA level differences. p300 is the major histone acetyltransferase that mediates H3K9 acetylation. ChIP assays showed that the

enrichment of p300 at the transgene loci was significantly higher for *opKRAS* cells than for the wild-type *KRAS* cells and was further increased for the *KRAS** cells (Figure 3.8E). These results indicate that the *KRAS* codon usage impacts transcription by affecting histone modifications and chromatin structure. Optimal codons result in transcriptionally permissive chromatin structures to promote recruitment of transcription co-activator, such as p300.

To determine whether the effect of codon usage on transcription is a general phenomenon or is *KRAS*-specific, I examined the effect of codon usage on *CFL2* expression. *CFL2* encodes an intracellular protein that is a major component of intranuclear and cytoplasmic actin rods. Mutation of *CFL2* results in human nemaline myopathy. Aside from its importance in human disease, I previously reported that *CFL2* is enriched in rare codons, and that changing rare codons to common increases the amount of ectopic *CFL2* protein (Lampson et al. 2013). I confirmed that *CFL2* protein expression is indeed greatly enhanced after codon optimization (Figure 3.9A). Similar to *KRAS*, codon optimization also led to a 6-fold increase of *CFL2* mRNA levels (Figure 3.9B). ChIP assays were performed to examine the enrichment of Pol II Ser 2 and Ser 5 phosphorylation and H3K9 acetylation at the transgene loci in cells that stably expressed either the wild-type *CFL2* or codon-optimized *CFL2* (Figure 3.9C). As expected, codon optimization resulted in a significant increase of enrichment of all three markers. Collectively, I conclude that codon usage also affects *KRAS* transcription and mRNA levels, an effect that may be applicable to other human genes enriched in rare codons.

The Differential Effects of Codon Usage on KRAS Expression in Different Cell Lines

Other lab had previously demonstrated that ectopic expression of KRAS* generated more protein than KRAS in a variety of cell types (Lampson et al. 2013; Pershing et al. 2015; Ali et al. 2017). However, the contribution of changing rare codons to common on *KRAS* mRNA levels in different cells, especially in light of the above results, had not been undertaken. To determine whether the effect of codon usage on *KRAS* mRNA levels is cell-line specific, I transfected the wild-type KRAS, opKRAS, and KRAS* expression constructs into two human hepatocellular carcinoma cell lines, *Huh7* and *HepG2*, and two human breast cancer cell lines MDA-MB-231 and MCF-7. Codon optimization resulted in increases of both KRAS protein and mRNA in each of these cell lines (Figures 3.10A-D). As in 293T cells, the fold changes of mRNA levels due to codon optimization in these cell lines were smaller than those of protein levels, suggesting that the effect of codon usage on translation efficiency is shared among these cell lines. However, it is clear that different cell lines responded differently to codon optimization. In HepG2 cells, the effects of codon usage on KRAS protein and RNA were much smaller than those in the other cell lines: Less than 10-fold more KRAS protein and less than 50% more mRNA were produced from *KRAS** than from wild-type *KRAS*. In contrast, in Huh7 cells the differences were larger than those seen in HEK-293T cells. In addition, the changes of *KRAS* transcript levels showed a strong positive correlation with the changes in protein levels in different cell lines (coefficient $r=0.93$, Figure 3.10E), indicating a major role for *KRAS* mRNA in determining KRAS protein levels. This suggest the intriguing possibility that the effect on codon bias at individual steps of protein production may be differentially regulated in different tissues.

Codon Optimization Alters KRAS Protein Structure

I and others have previously shown that codon usage affects the translation elongation rate in fungi and flies (Yu et al. 2015; Weinberg et al. 2016; Zhao et al. 2017), which in turn can influence protein structure during the co-translational folding process. Codon usage has been shown to regulate protein folding *in vitro* and in *E. coli*, fungi, and *Drosophila* cells (Komar et al. 1999; Spencer et al. 2012; Zhou et al. 2013; Sander et al. 2014; Yu et al. 2015; Zhou et al. 2015; Buhr et al. 2016; Fu et al. 2016; Zhao et al. 2017). To determine whether codon usage influences protein folding in mammals, I compared KRAS protein structures by performing a limited trypsin digestion assay, which can indicate protein structural differences. Extracts of 293T cells transfected with the wild-type KRAS or KRAS* expression constructs were used. In the presence of the same concentration of trypsin, KRAS protein expressed from the wild-type KRAS was much more resistant to trypsin digestion than that expressed from KRAS* (Figure 3.11A).

To further confirm our conclusion, I carried out thermal shift assays using extracts of 293T cells (Molina et al. 2013; Jafari et al. 2014). This assay quantifies changes in thermal denaturation and aggregation temperature of a protein as a result of treatment by increasing temperatures, which results in the disappearance of the protein from the supernatant. Changes in denaturation and aggregation temperature are indicative of structural changes of a protein. Increasing temperatures from 4-52.6°C resulted in a gradual slow disappearance of KRAS from the wild-type KRAS extract. In contrast, most of the KRAS from the KRAS* extract was precipitated and disappeared from the supernatant above 40°C (Figure 3.11B). Together, these results suggested that codon optimization of *KRAS* alters structural properties of KRAS proteins.

Summary

As described above, numerous experiments support the contention that the low level of KRAS is important for how the gene functions in normal and cancer biology of mammals, including in whole animal settings. As such, understanding how KRAS expression is maintained at a low level is critical. One feature of this gene that contributes to the poor expression of KRAS is due to its poor codon usage. Consistent with our previous observations, I show here that codon usage has a remarkable effect on KRAS expression from cDNA constructs: Codon optimization of *KRAS* resulted in the up-regulation of KRAS by about 100 fold (Figure 3.2 B). In addition, the effect of codon usage was accumulative. The number of codons optimized was correlated with levels of KRAS, and effects were not restricted to a specific region of *KRAS* open reading frame. These results suggest that the different codon usage profiles of *KRAS* and *HRAS* are the primary reason for their different protein levels. In agreement with our previous observations, this increase was attributed to increased translation and mRNA levels. At the translational level, optimal codons promote efficient translation of *KRAS* mRNA. This conclusion is supported by analyses of translation of *KRAS* mRNA constructs in cells and *in vitro* (Figures 3.5A-B) and by polysome profiling results that showed that codon optimization of *KRAS* led to the enrichment of the *KRAS* mRNA in the highly translated polysome fraction. Our lab has previously shown that optimal codons are known to increase the rate of translation elongation, and rare codons can result in ribosome stalling in fungi and fly cells (Yu et al. 2015; Weinberg et al. 2016; Zhao et al. 2017). Our data suggest that codon usage has a similar effect in human cells.

Codon usage also determines *KRAS* mRNA levels. However, unlike in yeast and some other organisms (Presnyak et al. 2015; Bazzini et al. 2016; Mishima and Tomari 2016; Radhakrishnan et al. 2016; Zhao et al. 2017), codon usage did not have a significant influence on *KRAS* mRNA stability in our experiments in human cells (Figure 3.7C). Instead, optimal codon usage promotes *KRAS* transcription. This conclusion is supported by nuclear run on and Pol II ChIP assay results (Figures 3.8A-B). In addition, I showed that codon optimization resulted in increases of H3K4me3 and H3K9ac, histone modifications that are associated with active chromatin (Figure 3.8D). Furthermore, I found that codon optimization led to enrichment of histone acetyltransferase p300 at the *KRAS* locus. Similar effects of codon optimization were also observed for *CFL2* gene (Figures 3.9 A-C), indicating that the transcriptional effect of codon usage may be a general phenomenon for human genes. Together, these results suggest that optimal codon usage affects transcription by recruiting co-transcriptional activators such as p300, which lead to chromatin modifications that alter chromatin structure and activate transcription.

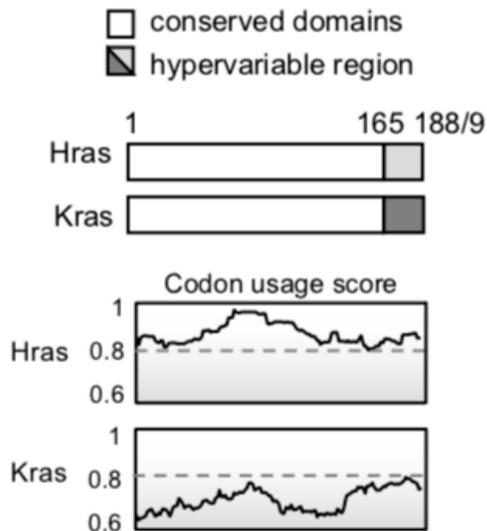
Consistent with our conclusion here, mammalian genes with high GC content, which is associated with more common codons, had higher expression levels than those with lower GC content without an effect on mRNA degradation rates (Kudla et al. 2006; Krinner et al. 2014). In addition, codon usage was shown to contribute to the balanced expression of Toll-like receptors in mammals through effects on transcription (Newman et al. 2016). Codon usage was previously shown to have a major role at the transcriptional level in *Neurospora* through regulation of chromatin structures (Zhou et al. 2016). Therefore, the role of codon usage on transcription appears to be conserved from fungi to human. How codon usage affects chromatin structures is

not known. Codon information within the open reading frame may be read by the transcription machinery in the form of DNA elements that favor or inhibit the recruitment of regulatory proteins that suppress or activate transcription.

Although a positive role for codon optimization on *KRAS* expression was observed in different cell lines, the degree of the codon usage effect differed (Figure 3.10). Such differential codon usage effects may be caused by different tRNA expression profiles in different cell lines (Dittmar et al. 2006), which are known to influence translation. In addition, the effects may be also due to differential expression levels of the chromatin regulatory factors that mediate the transcriptional effect of codon usage.

Finally, our results show that *KRAS* codon usage may also affect *KRAS* protein structure. We and others have previously shown that codon usage regulates translation elongation in fungi and *Drosophila* (Yu et al. 2015; Weinberg et al. 2016; Zhao et al. 2017). Changes in elongation rate change the time available for co-translational folding thus influencing protein structures. Consistent with this, codon usage has been shown to regulate protein folding *in vitro* and in *E. coli*, fungi, and *Drosophila* cells (Komar et al. 1999; Spencer et al. 2012; Zhou et al. 2013; Sander et al. 2014; Yu et al. 2015; Zhou et al. 2015; Buhr et al. 2016; Fu et al. 2016; Zhao et al. 2017). Our results suggest that this is also the case in human cells. Consistent with this conclusion, a single synonymous SNP that results in a rare codon in the human *MDR1* gene, which encodes a transporter, was found to result in altered drug and inhibitor interactions (Kimchi-Sarfaty et al. 2007). Furthermore, codon usage has been implicated in co-translational protein folding of CFTR, a protein that regulates transmembrane conductance, which is mutated in cystic fibrosis patients (Lazrak et al. 2013a; Kim et al. 2015; Kirchner et al. 2017). Taken

together, our results demonstrate that codon usage influences gene expression and protein structure in human cells by multiple mechanisms. Because many human diseases are known to be associated with synonymous mutations (Sauna and Kimchi-Sarfaty 2011; Birkeland et al. 2012; McCarthy et al. 2017), our study suggests how these mutations can contribute to disease progression without altering amino acid sequences.

A**B**

	Hras	Kras	
Valine (V)	GTG	100%	13.33%
	GTC	0	0
	GTT	0	33.33%
	GTA	0	53.33%
Isoleucine (I)	ATC	83.33%	0
	ATT	8.33%	75%
	ATA	8.33%	25%

C

Figure 3.1 Codon usage differences mediates differential protein expression of Hras and Kras. A) Top panels: Schematic diagram of the sequence of Hras and Kras cDNA sequences. Bottom panels: codon usage blot of Hras and Kras along the sequence. B) Hras and Kras use very different codons to decode amino acids Valine and Isoleucine. C) Western blot results showing the levels of HRAS and KRAS in HEK-293 cells transiently transfected with cDNA expression construct.

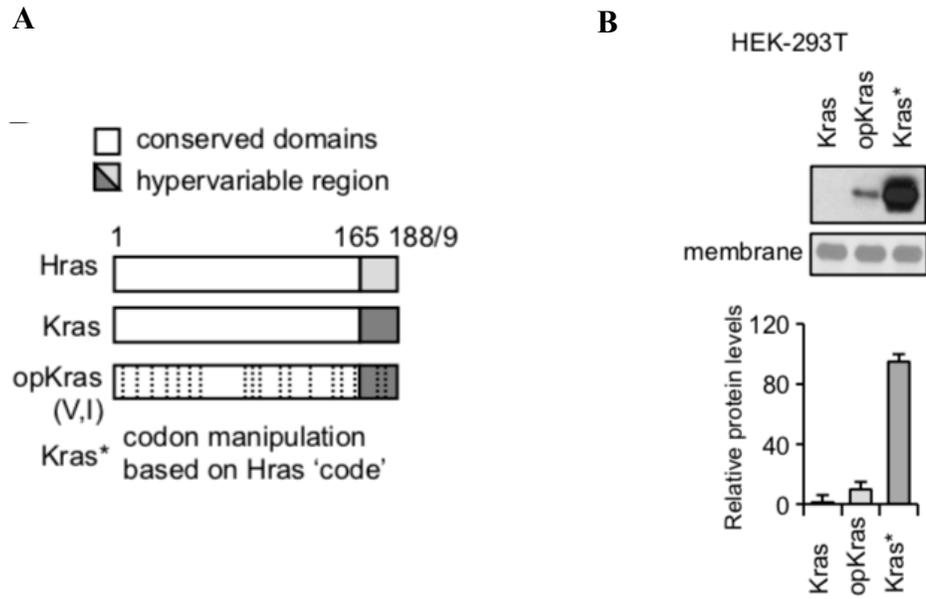
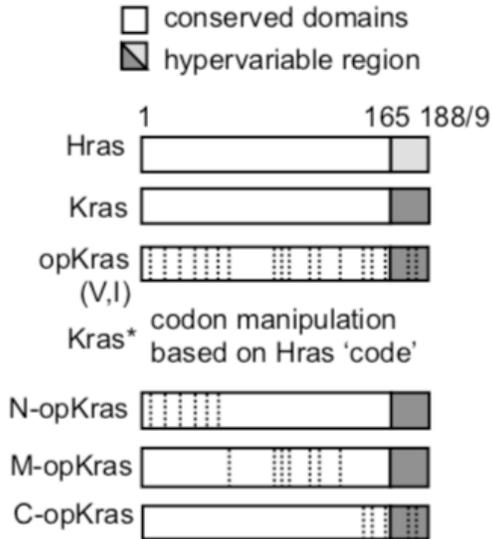


Figure 3.2 Rare codons suppress Kras protein expression.

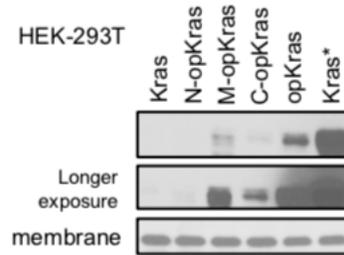
A) Schematic diagram of different versions of *KRAS* ORF inserted in the expression construct. All constructs share the same promoter, 5' and 3' UTR. Red lines marked the codon positions that were mutated.

B) Top panels: Western blot results showing protein level of KRAS in HEK-293 cells transfected with the indicated expression construct. Bottom: densitometric analyses of the Western blot results from three independent experiments. The levels of Kras were set as 1. Error bars indicate \pm SD.

A



B



C

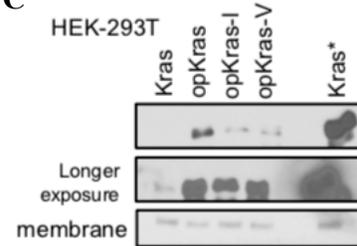


Figure 3.3 Codon composition determines the protein expression levels of Kras.

A) Schematic diagram of different versions of *KRAS* ORF inserted in the expression construct. All constructs share the same promoter, 5' and 3' UTR. Red lines marked the codon positions that were mutated.

B) Accumulative effects of codon on protein expression levels.

C) Western blot results showing only optimization of rare codons of Valine (opKras-V) and Isoleucine (opKras-I) could enhance protein expressions.

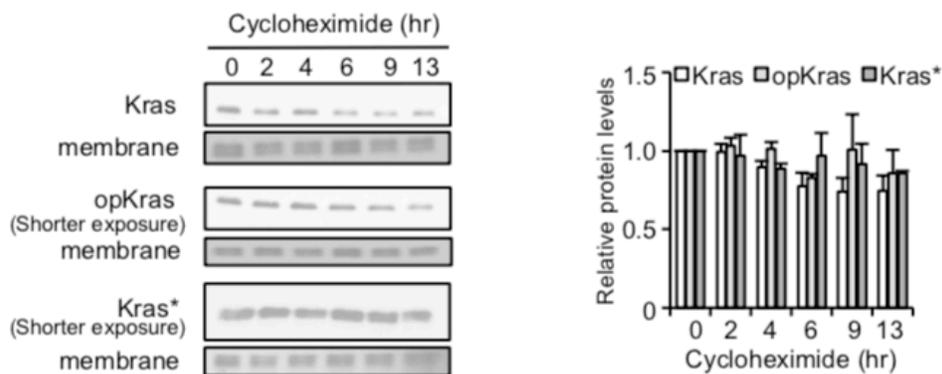


Figure 3.4 Codon optimization does not change protein stability of Kras.

Cells were treated with cycloheximide (CHX), and harvested at the indicated time points. Left panel: western blot results showing different codon-optimized versions of Kras protein degraded in a similar speed; Right panel: densitometric analyses of the western blot results from three independent experiments. Error bars denote \pm s.d.

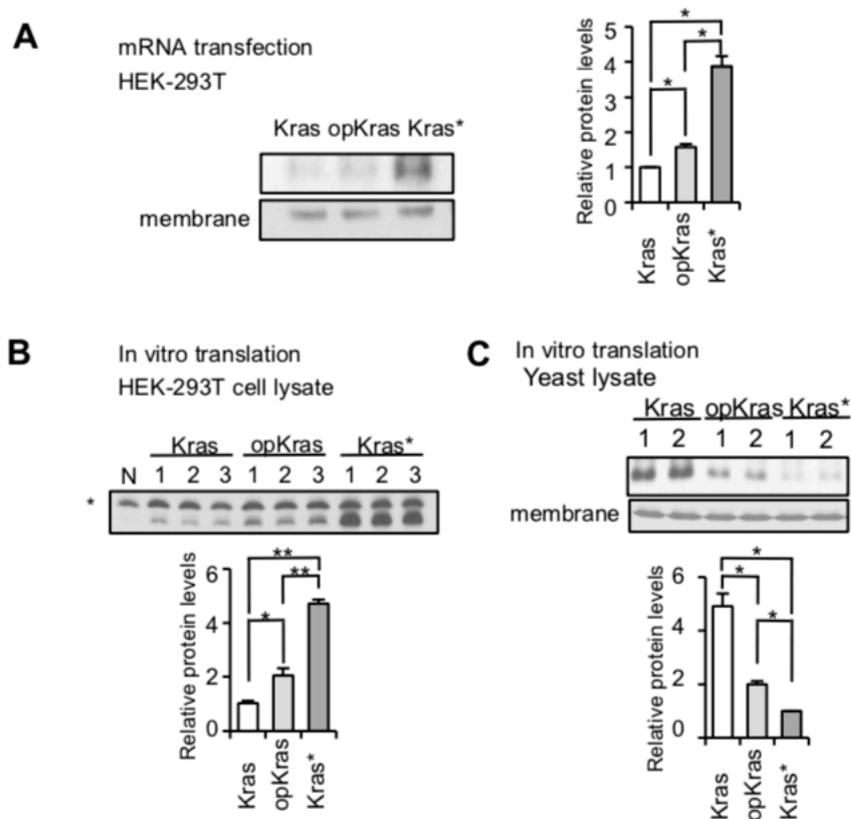


Figure 3.5 Codon optimization promotes translation of KRAS mRNA.

A) Left panel: HEK-293T cells were transfected with same amount of in-vitro synthesized *KRAS* mRNAs, and the levels of KRAS produced were measured by western blot. Right panel: densitometric analyses of the western blot results from three independent experiments.

B) Top panel: western blot results showing the levels of KRAS produced from the indicated in-vitro synthesized *KRAS* mRNA in HEK-293 cell translation extracts. The asterisk indicates a non-specific protein band. Bottom panel: densitometric analyses of the western blot results from three independent experiments.

C) Top panel: in-vitro translation assay results showing relative protein levels of KRAS produced the indicated *KRAS* mRNA in yeast translation extracts. Bottom panel: densitometric analyses of the western blot results from three independent experiments.

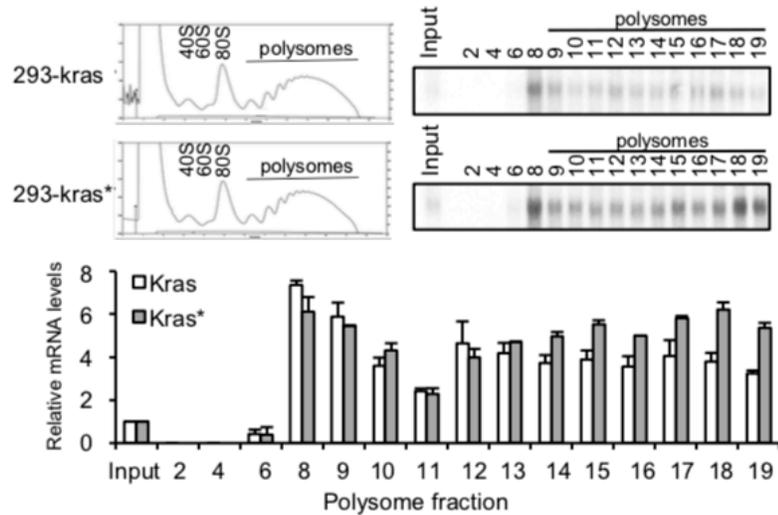


Figure 3.6 Codon-optimized Kras mRNAs were associated with active translated polysomes.

Top left panels: representative absorbance profile for polysome gradient from 293T-Kras and 293T-Kras* stable cell lines. The fraction numbers of the 40S, 60S, 80S ribosome, and polysome are indicated. Top right panels: northern blot results showing the levels of KRAS mRNA in the polysome fractions. Bottom panel: densitometric analyses of the northern blot results in different polysome fractions. The mRNA levels in each fraction were normalized by input DNA. Error bars denote \pm s.d. * $P < 0.01$; ** $P < 0.001$.

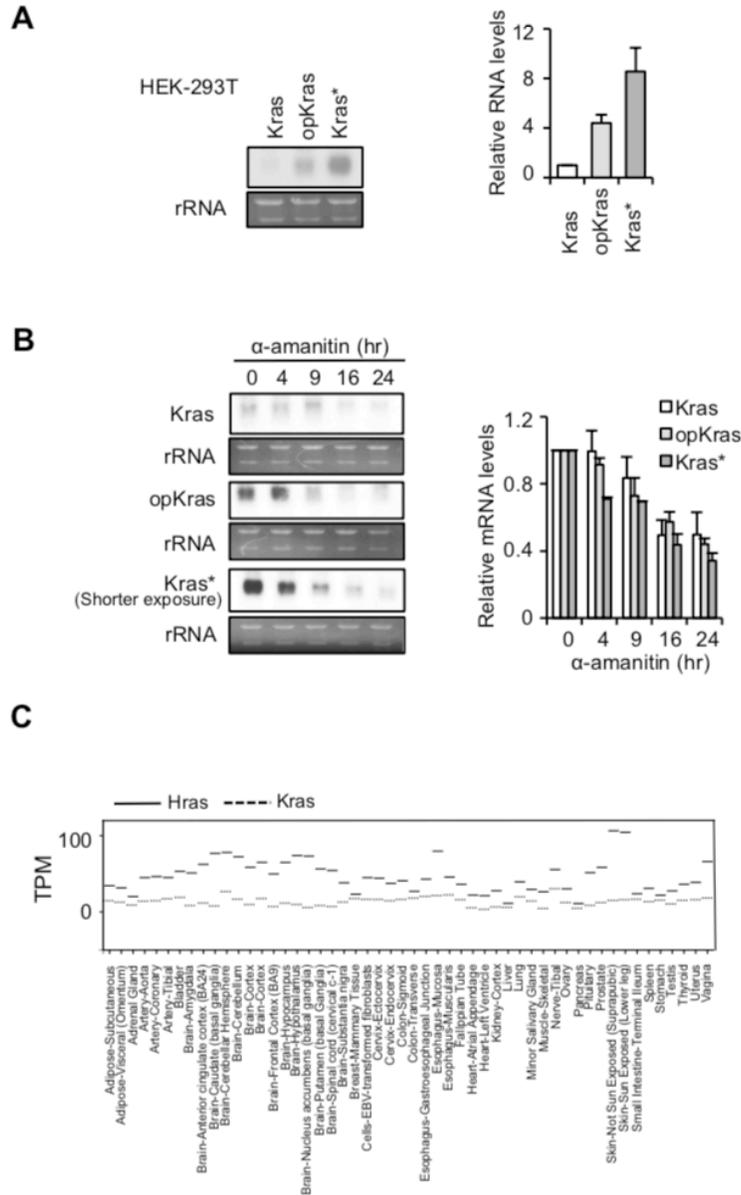


Figure 3.7 Codon optimization increases KRAS mRNA levels without affecting mRNA stability.

A) Northern blot showing the *KRAS* mRNA levels in HEK-293 cells transfected with the indicated *Kras* construct.

B) Northern blot analysis showing the decay of *KRAS* mRNA after the addition of α -amanitin. Right panel: densitometric analyses of the northern blot results from three independent experiments. Error bars denote \pm s.d.

C) Comparison of *KRAS* and *HRAS* mRNA levels in different human tissues. The results were obtained from the Genotype-Tissue Expression (*GTEx*) Program (<https://gtexportal.org/home/>)

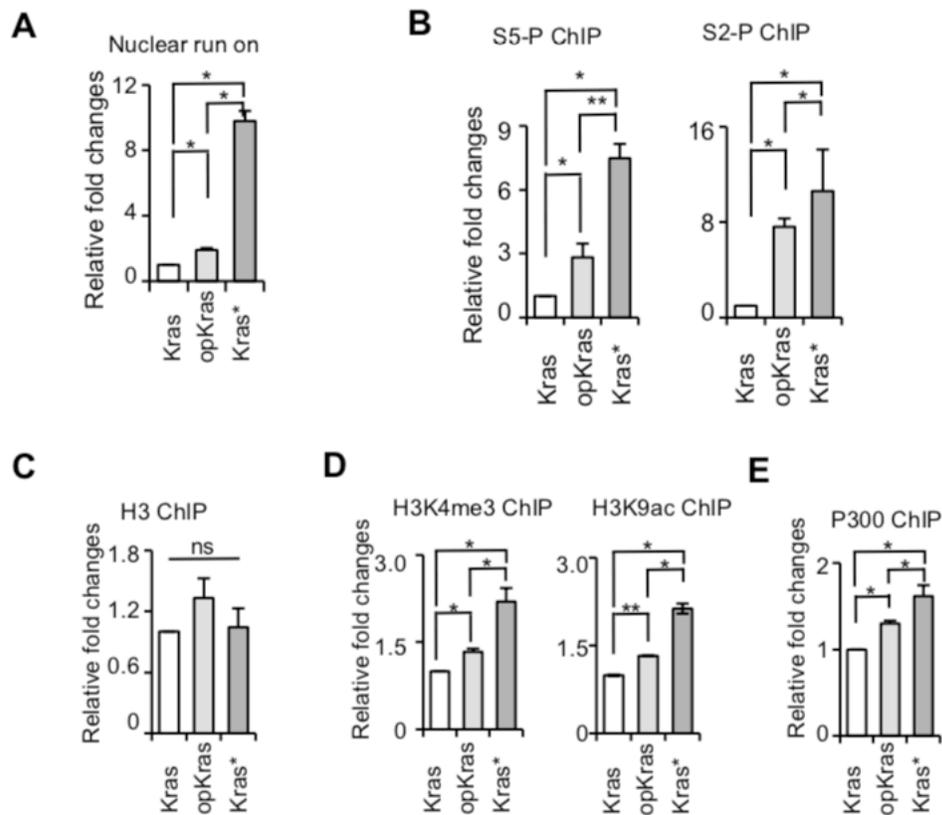


Figure 3.8 Codon optimization promotes KRAS transcription.

A) Nuclear run on assay results showing the relative levels of nascent KRAS mRNA in the indicated cell lines. HEK-293 cells stably transfected with the indicated Kras construct were used.

B) ChIP assay results showing the relative enrichment of phosphorylated Ser-5 (S5P) at the promoter region and phosphorylated Ser-2 (S2P) at the 3' UTR of the Kras transgene. The ChIP results were first normalized by input DNA and then normalized by the wild-type Kras signal.

C) ChIP assay results showing the levels of histone H3 were not significantly altered by KRAS codon optimization.

D) H3K4me3 and H3K9ac ChIP assays showing the relative histone modification levels at the indicated Kras transgene loci.

E) p300 ChIP assay results showing the enrichment of p300 proteins at the promoter regions of the indicated Kras transgenes. Error bars denote \pm s.d.

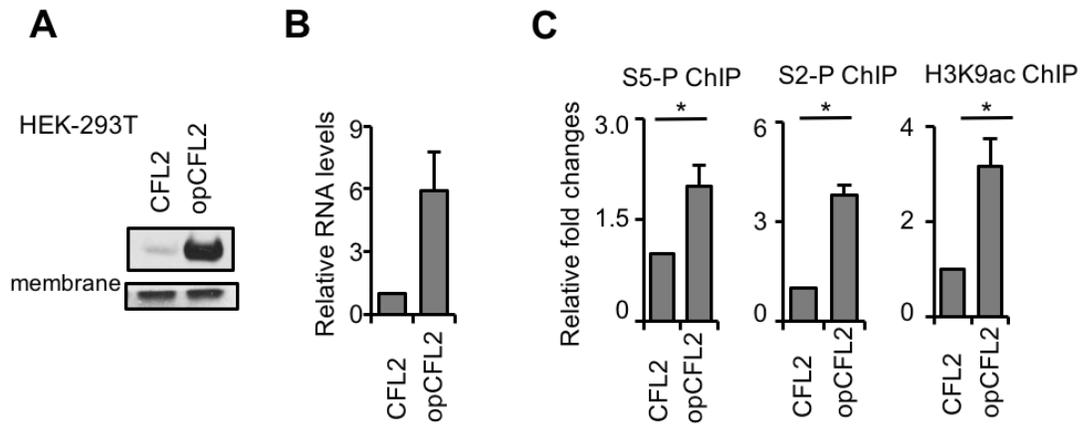


Figure 3.9 Codon effect on transcription is a general phenomenon.

A) Western blot results showing CFL2 protein level in HEK-293 cells stably transfected with the indicated CFL2 expression construct.

B) Quantitative RT-PCR result showing the transcript levels of CFL2.

C) ChIP assay results showing the relative enrichment of phosphorylated Ser-5 (S5P) at the promoter region and phosphorylated Ser-2 (S2P) at the 3' UTR of the indicated CLF2 transgene.

* $P < 0.01$; ** $P < 0.001$.

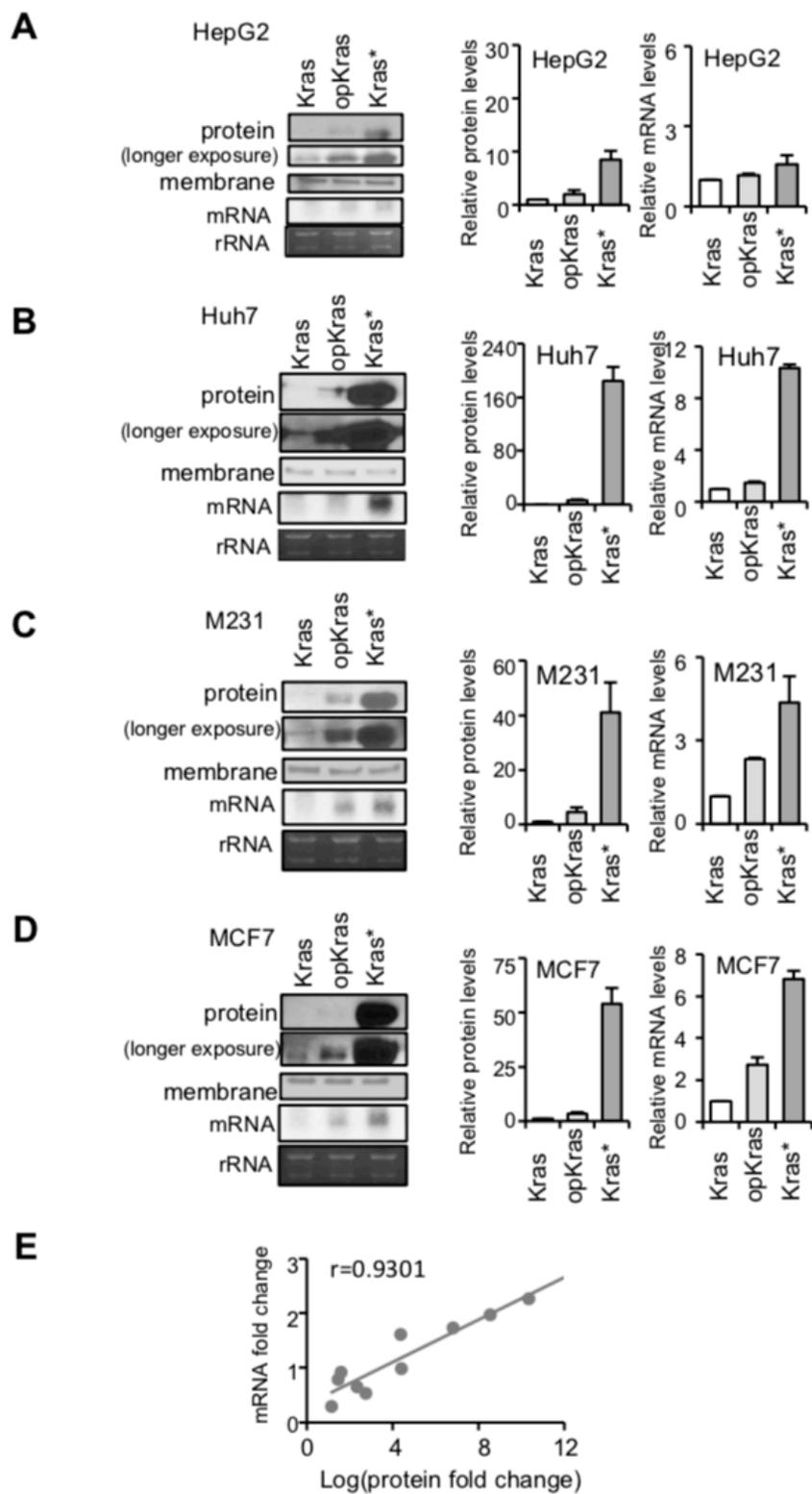


Figure 3.10 Differential KRAS codon usage responses in different cell lines.

A-D) Left panels: western and northern blot results showing the levels of KRAS protein and mRNA levels in HepG2, Huh7, M231, and MCF7 cell lines transiently transfected with the indicated Kras expression construct. Right panel, densitometric analyses of the western and northern blot results for each cell line. Error bars denote \pm s.d.

E) Scatter analysis showing the correlation between mRNA fold changes and log (protein fold changes). Pearson's $r = 0.93$.

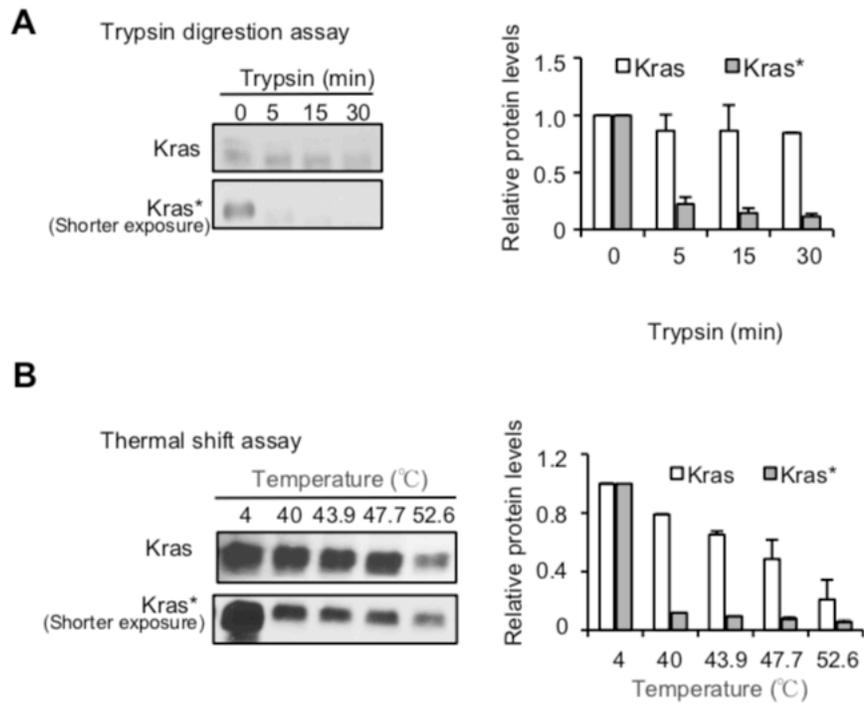


Figure 3.11 Codon optimization results in altered KRAS structure.

A) Left panel: western blot results showing the KRAS protein levels after partial trypsin (0.5 mg/ml) digestion at the indicated time points. HEK-293 cell extracts from cells transiently transfected with the Kras and Kras* construct were used. Right panel: densitometric analyses of the western blot results from three independent experiments.

B) Thermal shift assay. Left panel: western blot results showing the KRAS levels in the supernatant after thermal treatment at the indicated temperatures. Right panel: densitometric analyses of the western blot results from three independent experiments.

opKras ATGGACTACAAAGACGATGACGACAAGACTGAATATAAACTTGTGGTGGT
N-opKras ATGGACTACAAAGACGATGACGACAAGACTGAATATAAACTTGTGGTGGT
M-opKras ATGGACTACAAAGACGATGACGACAAGACTGAATATAAACTTGTGGTGGT
C-opKras ATGGACTACAAAGACGATGACGACAAGACTGAATATAAACTTGTGGTGGT
opKrasI ATGGACTACAAAGACGATGACGACAAGACTGAATATAAACTTGTGGTGGT
opKrasV ATGGACTACAAAGACGATGACGACAAGACTGAATATAAACTTGTGGTGGT
***** **

opKras TGGAGCTGGTGGCGTGGGCAAGAGTGCCTTGACGATACAGCTAATCCAGA
N-opKras TGGAGCTGGTGGCGTGGGCAAGAGTGCCTTGACGATACAGCTAATCCAGA
M-opKras TGGAGCAGGAGGAGTAGGCAAGAGTGCCTTGACGATACAGCTAATCCAGA
C-opKras TGGAGCAGGAGGAGTAGGCAAGAGTGCCTTGACGATACAGCTAATCCAGA
opKrasI TGGAGCAGGAGGAGTAGGCAAGAGTGCCTTGACGATACAGCTAATCCAGA
opKrasV TGGAGCAGGAGGAGTAGGCAAGAGTGCCTTGACGATACAGCTAATCCAGA
***** ** ** * *****

opKras ATCATT TTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAG
N-opKras ATCATT TTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAG
M-opKras ATCATT TTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAG
C-opKras ATCATT TTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAG
opKrasI ATCATT TTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAG
opKrasV ATCATT TTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAG

opKras CAAGTGGTGATCGATGGAGAAACCTGTCTCTTGGATATCCTCGACACAGC
N-opKras CAAGTGGTGATCGATGGAGAAACCTGTCTCTTGGATATCCTCGACACAGC
M-opKras CAAGTAGTAATTGATGGAGAAACCTGTCTCTTGGATATCCTCGACACAGC
C-opKras CAAGTAGTAATTGATGGAGAAACCTGTCTCTTGGATATCCTCGACACAGC
opKrasI CAAGTAGTAATCGATGGAGAAACCTGTCTCTTGGATATCCTCGACACAGC
opKrasV CAAGTGGTGATCGATGGAGAAACCTGTCTCTTGGATATCCTCGACACAGC
***** ** ** *****

opKras AGGTCAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGG
N-opKras AGGTCAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGG
M-opKras AGGTCAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGG
C-opKras AGGTCAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGG
opKrasI AGGTCAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGG
opKrasV AGGTCAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGG

opKras AGGGCTTTCTTTGTGTGTTTGGCATAAATAACTAAATCATT TGAAGAT
N-opKras AGGGCTTTCTTTGTGTATTTGGCATAAATAACTAAATCATT TGAAGAT
M-opKras AGGGCTTTCTTTGTGTGTTTGGCATAAATAACTAAATCATT TGAAGAT
C-opKras AGGGCTTTCTTTGTGTATTTGGCATAAATAACTAAATCATT TGAAGAT
opKrasI AGGGCTTTCTTTGTGTATTTGGCATAAATAACTAAATCATT TGAAGAT
opKrasV AGGGCTTTCTTTGTGTGTTTGGCATAAATAACTAAATCATT TGAAGAT

opKras	ATCCACCATTATAGAGAACAAATCAAAGAGTTAAGGACTCTGAAGATGT
N-opKras	ATTCACCATTATAGAGAACAAATTTAAAAGAGTTAAGGACTCTGAAGATGT
M-opKras	ATCCACCATTATAGAGAACAAATCAAAGAGTTAAGGACTCTGAAGATGT
C-opKras	ATTCACCATTATAGAGAACAAATTTAAAAGAGTTAAGGACTCTGAAGATGT
opKrasI	ATCCACCATTATAGAGAACAAATCAAAGAGTTAAGGACTCTGAAGATGT
opKrasV	ATTCACCATTATAGAGAACAAATTTAAAAGAGTTAAGGACTCTGAAGATGT
	** *****
opKras	GCCTATGGTCCTAGTGGGAAATAAATGTGATTTGCCTTCTAGAACAGTGG
N-opKras	ACCTATGGTCCTAGTAGGAAATAAATGTGATTTGCCTTCTAGAACAGTAG
M-opKras	GCCTATGGTCCTAGTGGGAAATAAATGTGATTTGCCTTCTAGAACAGTAG
C-opKras	ACCTATGGTCCTAGTAGGAAATAAATGTGATTTGCCTTCTAGAACAGTGG
opKrasI	ACCTATGGTCCTAGTAGGAAATAAATGTGATTTGCCTTCTAGAACAGTAG
opKrasV	GCCTATGGTCCTAGTGGGAAATAAATGTGATTTGCCTTCTAGAACAGTGG
	***** *
opKras	ACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATCCCTTTTATC
N-opKras	ACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATCCCTTTTATT
M-opKras	ACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATCCCTTTTATT
C-opKras	ACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATCCCTTTTATC
opKrasI	ACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATCCCTTTTATC
opKrasV	ACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATCCCTTTTATT

opKras	GAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATT
N-opKras	GAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATT
M-opKras	GAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATT
C-opKras	GAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATT
opKrasI	GAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATT
opKrasV	GAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATT

opKras	AGTTCGAGAAATCCGAAAACATAAAGAAAAGATGAGCAAAGATGGTAAAA
N-opKras	AGTTCGAGAAATTCGAAAACATAAAGAAAAGATGAGCAAAGACGGTAAAA
M-opKras	AGTTCGAGAAATTCGAAAACATAAAGAAAAGATGAGCAAAGACGGTAAAA
C-opKras	AGTTCGAGAAATCCGAAAACATAAAGAAAAGATGAGCAAAGATGGTAAAA
opKrasI	AGTTCGAGAAATCCGAAAACATAAAGAAAAGATGAGCAAAGACGGTAAAA
opKrasV	AGTTCGAGAAATTCGAAAACATAAAGAAAAGATGAGCAAAGACGGTAAAA

opKras	AGAAGAAAAAGAAGTCAAAGACAAAGTGTGTGATCATGTAA
N-opKras	AGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAATTATGTAA
M-opKras	AGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAATTATGTAA
C-opKras	AGAAGAAAAAGAAGTCAAAGACAAAGTGTGTGATCATGTAA
opKrasI	AGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAATTATGTAA
opKrasV	AGAAGAAAAAGAAGTCAAAGACAAAGTGTGTGATTATGTAA
	***** ** *****

Table 3.1 Based on pBabe-opKras plasmid, pBabe-opKrasV, -opKrasI, -N-opKras, -M-opKras, and -C-opKras plasmids were designed, and the sequence were shown.

CHAPTER FOUR

MATERIAL AND METHODS

Calculation of the CAI and CBI

CAI was calculated as previously described (Sharp and Li 1987). Codon usage frequency table for *D. melanogaster* was obtained from <http://www.kazusa.or.jp/codon/>. CBI was calculated by codonw (<http://mobyli.pasteur.fr/cgi-bin/portal.py#forms::codonw>) (Bennetzen and Hall 1982).

Calculation of Protein Structural Disorder Tendency

IUPRED (<http://iupred.enzim.hu/>) was used to predict protein disorder tendency.

Constructs used in Fly system

A previously described CaSpeR-4-based transformation vector, 13.2(per⁺-HA10His), was used as the parent vector for codon optimization (Lee et al. 1998). This vector contains a 13.2 kb genomic dper region, in which dPER protein was tagged by a HA epitope tag and a stretch of histidine residues at the C-terminus. Codon-optimized sequences encoding amino acids 15-21; 43-316 (OP1) and 539-982 (OP2) were synthesized by Genscript and cloned into the above mentioned vector to yield 13.2(per(op1)-HA10His) and 13.2(per(op2)-HA10His).

Generation of Transgenic Flies

Transgenic flies were generated by BestGene Inc. using standard P element-mediated transformation techniques using w¹¹¹⁸ embryos. At least three independent germ-line

transformants bearing the *dper*-containing plasmids were obtained and then crossed into a *wper*⁰ genetic background to yield *wper*⁰; *p*{*dper*(OP1)} and *wper*⁰; *p*{*dper*(OP2)} (referred as *p*{*dper*(OP1)} and *p*{*dper*(OP2)}). Transgenic flies carrying 13.2(*per*⁺-HA10His) were used as control (*wper*⁰; *p*{*dper*(WT)}). The *p*{*dper*(OX)} flies are flies with *w*¹¹¹⁸ genetic background carrying extra copies of the wild-type version of the *dper* transgene.

Drosophila locomotor activity analysis

The locomotor activities of individual flies were measured as previously described using the monitoring system from Trikinetics (Chiu et al. 2010). Briefly, 1~5 d adult male flies were used for the analysis and kept in incubators at 25 °C, exposed to 4 d of 12 h light followed by 12 h dark [12:12 LD, where zeitgeber time 0 (ZT0) is defined at the time when the light phase begins and subsequently kept in constant darkness for 5-8 d to measure free-running rhythm. The locomotor activity data for each individual fly was analyzed using the FaasX software, which was generously provided by F. Rouyer (Centre National de la Recherche Scientifique, Paris, France). Periods were calculated for each individual fly using periodogram analysis and pooled to obtain a group average for each independent transgenic line or genotype. Power is a measure of the relative strength of the rhythm during DD. individual flies with a power ≥ 10 and a “width” value of 2 or more (denotes number of peaks in 30 min increments above the periodogram 95% confidence line) were considered rhythmic.

Quantification of morning and evening anticipations in LD cycles were calculated using the method described previously (Cusumano et al. 2009). An anticipation index (AI) was obtained as the slope of a linear regression through the last eight 30-minute bins before Lights-

OFF (evening anticipation) or the last five bins before Lights-ON (morning anticipation). It thus represents an average variation in activity counts / 30 minute-bin, from one bin to the next within that time-window.

Immunoblotting

Flies were collected by freezing at the indicated times in LD or DD and total head extracts prepared using either EB1 (20mM Hepes pH 7.5, 100mM KCl, 5% glycerol, 2.5mM EDTA, 5mM DTT, 0.1% Triton X-100, 0.5mM PMSF, 10ug/ml Aprotinin, 5ug/ml Leupeptin, 1ug/ml Pepstatin A, 25mM NaF), or modified-RIPA buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 1mM EDTA, 0.25% sodium deoxycholate, 1% NP-40, 0.5mM PMSF, 25mM NaF, Roche complete protease inhibitor (EDTA-free), Roche PhosStop phosphatase inhibitor) with sonication depending on which proteins i sought to detect; i.e., EB1 buffer was used for dPER and TIM, whereas m-RIPA with sonication was used for CLK (Figure 3B). In the case of S2 cell extracts, the cells were harvested at the indicated times after kinase induction (Figure 6B, 6C) and lysed using EB2 lysis buffer (20mM Hepes pH 7.5, 100mM KCl, 5% glycerol, 5mM EDTA, 1mM DTT, 0.1% Triton X-100, 0.5mM PMSF, 10ug/ml Aprotinin, 5ug/ml Leupeptin, 1ug/ml Pepstatin A, 25mM NaF). Primary antibodies were used at the following dilutions: anti-V5 (Invitrogen), 1:7000; anti-mouse HRP (GE), 1:2000; anti-HA 3F10 (Roche) 1:1000 for Fly extracts; anti-rat HRP (GE), 1:1000; anti-Per (GP5620) 1:3000; anti-GP HRP (Sigma), 1:2000; anti-Tim (R1) 1:1000; anti-GP HRP (Sigma), 1:2000; anti-goat HRP (Santa Cruz), 1:5000; anti-CLK (Santa Cruz): 1:1000; anti-goat HRP (Santa Cruz), 1:1000. SDS-PAGE Gels (6%) were used to resolve dPER and TIM, and in the case of CLK, 5% Tris-HCl Criterion gels (Bio-Rad)

were used. All western blots were imaged using the Chemidoc software for the Biorad Chemidoc, which include correction for background signals.

For mammalian protein analysis, cells were lysed in RIPA (25 mM Tris, pH 7.4, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 1 mM DTT, 0.5 mM PMSF, 10 µg/ml aprotinin, 5 µg/ml leupeptin, 1 µg/ml pepstatin A, 25 mM NaF). Primary and secondary antibodies were used at the following dilutions: anti-FLAG, 1:2000 (Sigma); anti-myc, 1:5000 (Invitrogen); and anti-mouse HRP, 1:2000 (Bio-Rad). SDS PAGE gels (12.5%) were used to resolve RAS proteins. For protein stability assays, cells were grown and transfected with indicated plasmids for 2 days before the addition of CHX (final concentration 10 µg/ml) and were collected at the indicated time points.

Immunohistochemistry

Confocal imaging of adult brains was performed as described previously (Ko et al. 2007). Briefly, adult flies were entrained and dissected from each ZT time point and incubated briefly in cold isopropyl alcohol. Fly heads were removed in an embryo dish (Electron Microscopy Sciences) filled with PBS and then transferred to a 1% collagenase solution for 10 minutes with agitation. Collagenase was removed and fixative (4% formaldehyde, 0.2% triton X-100, in PBS) was added. Heads were fixed for 45 minutes at RT with agitation. Fixative was then removed and heads were rinsed twice, then washed twice for 30 minutes. Heads were transferred to blocking solution (0.2% triton X-100, 5% normal goat or horse serum, in PBS). Brains were dissected using #5 Rubis nano tweezers (Electron Microscopy Sciences) in an embryo dish filled with blocking solution. Brains were blocked between 1-2 hrs and were incubated overnight with

primary antibodies, anti-HA (clone 3F10, Roche, Indianapolis, IN) at 1:750 and anti-PDF (DSHB, Iowa city, IA) at 1:200 for all brains. Brains were incubated in primary antibodies overnight (O/N) in blocking at 4°C with agitation. After ~16 hrs, brains were rinsed twice and then washed twice for 30 minutes in wash solution (0.2% triton X-100 in PBS). Then secondary antibodies, which were conjugated to Alexa dyes (Life Technologies, Grand Island, NY) for visualization on a confocal microscope, were added to brains in blocking solution. Brains were incubated with secondary antibodies: Alexa Fluor 488-conjugated goat anti-rat (against anti-HA) and Alexa Fluor 594-conjugated goat anti-mouse (against anti-PDF) at 4°C for 4 hrs with agitation. After secondary antibody incubation, brains were once again rinsed twice and washed twice for 30 minutes in wash solution. Finally, brains were mounted upon microscope slides in VectaShield mounting media (Vector labs, Burlingame, CA) under a #1.5 (17mm) cover slip. Prepared slides were stored in 4°C in a light blocking container to prevent bleaching of fluorescent dye.

Quantitative real-time PCR

Total RNA was isolated from frozen heads using TRIzol (Invitrogen). Five hundred nanograms of total RNA were reverse transcribed using a High-capacity cDNA Reverse transcription Kit (ABI) in accordance with the manufacturer's instructions and subjected to real-time PCR analysis. Primer sequences used here for quantitation are as follows: dper forward: 5'-CGCAGCATCATGGACTTCTA-3'; dper reverse: 5'-CCGTCTGACCCTTCTTCATTAC-3'; tim forward: 5'-CCTTTTCGTACACAGA-TGCCA-3'; tim reverse: 5'-GGTCCGTCTGGTGATCCCAG-3' (Kadener et al. 2007); dcwo forward: 5'-GTCTGTGGA-

TCGAGGAGCAG-3'; dcwo reverse: 5'-GGCATATTCAGCATCGTCCT-3' (Kadener et al. 2007); dgol forward: 5'-GCCACGGATCTATGCAGTTT-3'; dgol reverse: 5'-CTTGGATAGCGACTGC-TGTG-3' (Abruzzi et al. 2011).

Immunoprecipitation assays

For immunoprecipitation, total head extracts were prepared using a modified-RIPA buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 1mM EDTA, 0.25% sodium deoxycholate, 1% NP-40, 0.5mM PMSF, 25mM NaF, Roche complete protease inhibitor (EDTA-free)) with sonication. 20 µl of either HA-agarose (Sigma) was added to total fly head extracts. Immune complexes were incubated with gentle rotation for 6 hr at 4°C. Beads were collected and mixed with 2× SDS sample buffer. 6% Gels were used to resolve dPER, and 5% Tris-HCl Criterion gels (Bio-Rad) were used to detect CLK.

To detect signals with the phospho-specific antibodies, Roche PhosStop phosphatase inhibitor was added to M-RIPA buffer. To immunoprecipitate dPER, 30 µl of anti-HA-Agarose beads (A2095, Sigma) was added to the extracts, and incubated with gentle rotation for 4 hr at 4°C. Proteins were eluted with SDS sample buffer and resolved by 6% SDS-PAGE. To perform Lambda phosphatase treatment, 0.6ul Lambda phosphatase (NEB) was added after dPER immunoprecipitation, and incubated at 30°C for 30 minutes. Proteins were eluted with SDS sample buffer and resolved by 6% SDS-PAGE.

Trypsin sensitivity assay

Protein concentration of the extract was diluted to 2.5 $\mu\text{g}/\mu\text{l}$. A 100- μl aliquot of extract was treated with trypsin (final concentration was experimentally determined) at room temperature with gentle shaking. A 20- μl sample was taken from the reaction at each time point (0, 5, 15, and 30 min) after addition of trypsin. Each 20- μl sample was mixed with protein loading buffer, and proteins were resolved on an SDS-PAGE gel (7.5%). Western blot analysis was performed to examine dPER protein levels at each time point (Zhou et al. 2013). Assays for different extracts were performed side by side and protein samples were transferred to the same membrane for western blot analysis.

Thermal shift assay

Flies were collected by freezing at ZT20 and total head extracts were prepared using M-RIPA buffer. Aliquots of protein lysates (18 μl) were heated at different temperatures for 2 mins (C1000 Thermal Cycler PCR machine, BioRad) followed by cooling for 3 min at room temperature. The lysates were then centrifuged at 15000 g for 20 minutes at 4°C to separate the soluble fractions from precipitates. The amount of dPER protein in the supernatants and precipitates were then analyzed by western blot analysis.

S2 culture and transfection

The pAct-dper-V5 and pMT-dbt-V5 plasmids were described previously (Kim et al. 2007). pAct-dper(op1)-V5 was generated by replacing wild-type dper sequence in the pAct-dper-V5 with OP1 sequence. All constructs were verified by DNA sequencing. S2 cells were transfected using effectene (Qiagen) following the manufacturer's protocol (Ko et al. 2002; Chiu

et al. 2008). For each transient transfection, 0.8 µg of different dper-containing plasmids and 0.2 µg of pMT-dbt-V5 plasmids were used. Expression of dbt under pMT promoter was induced by adding 500 µM CuSO₄ to the culture media for 36 hr after transfection. For experiments in which the proteasome inhibitor MG132 (50 µM; Sigma) and cycloheximide (10 µg/ml; Sigma) were used, they were added 4 hr prior to cell harvest. To quantify different phosphorylation isoforms, ImageJ is used to plot for the relative attribution of signals along each loading lane. To quantify the relative ratio of phosphorylation isoforms in the in vitro degradation assay, percentage of top half (hyper-phosphorylated) or bottom half (hypophosphorylated) is calculated over total signal.

GST-pull down assay

To generate GST-DBT proteins, expression was induced from stable cell lines for 36 h, and cells were lysed in GST lysis buffer (20mM Tris-HCl pH 7.5, 0.05% IGEPAL CA-60, 1mM EDTA, 5mM DTT, 150mM NaCl, 25mM NaF, complete EDTA free protease inhibitor cocktail (Roche)). The extracts were then incubated with glutathione beads overnight at 4°C to achieve binding. Roughly 250µl of heads were used to prepare extracts for each GST pull-down reaction. Heads were homogenized in Modified RIPA buffer (see methods for IP) with the addition of complete EDTA-free protease inhibitor cocktail (Roche). Prior to the pull-down, proteins were quantified using spectrophotometer to ensure that equal amounts of head extracts were used for each reaction. Bound dPER was resolved using 6% SDS-PAGE to determine dPER-DBT interaction.

Mammalian Cell Lines

HEK-293T, HepG2, MDA-MB-231, and MCF7 cells were maintained at 37 °C in 5% CO₂ in DMEM (Sigma) supplemented with 10% fetal bovine serum (Sigma) and 100 U/ml penicillin and streptomycin. Huh7 cells were maintained in McCoy's 5A media (Gibco/Invitrogen) with identical supplements. For expression studies, cells were either transiently transfected with plasmids using PolyJet (SignaGen) according to the manufacturer's instructions or selected for stable expression of constructs by puromycin or neomycin antibiotic resistance.

Plasmids

pBabe-Kras, -opKras, and -Kras* expression constructs were created previously (Lampson et al. 2013). The other constructs used in this study were created based on these plasmids.

RNA analysis

RNA was extracted with Trizol (Ambion) in accordance with the manufacturer's protocol. For northern blot analyses, equal amounts of total RNA (5 µg) were loaded onto agarose gels. After electrophoresis, the RNA was transferred onto nitrocellulose membrane (GVS North America). The membrane was probed with an RNA probe specific for 5' untranslated region of the KRAS mRNAs. The probe was labeled with [³²P] UTP (PerkinElmer) during transcription by T7 RNA polymerase (Ambion) with the manufacturer's protocol. The primer sequences used for the template amplification were Northern forward: 5'-

CCTTAGGTCACTGGAAAGATG-3', and Northern reverse: 5'-

TAATACGACTCACTATAGGGGTCGTCATCGTCTTTGTAGTC-3'.

For quantitative real-time PCR analysis, 500 ng of total RNA were reverse transcribed using a High-capacity cDNA Reverse Transcription Kit (ABI) in accordance with the manufacturer's instructions and subjected to real-time PCR analysis. Primer sequences used for quantitation are

as follows: ras forward: 5'- AGCCCTTTGTACACCCTAA-3', ras reverse: 5'-

GTCGTCATCGTCTTTGTAGTC-3'; gapdh forward: 5'

CATGTTTCGTCATGGGTGTGAACCA-3', gapdh reverse: 5'-

AGTGATGGCATGGACTGTGGTCAT-3'.

For the RNA stability assay, cells were grown and transfected with indicated plasmids for 2 days before the addition of α -amanitin (final concentration 50 μ g/ml), and collected at the indicated time points.

Nuclear run on assay

Cells were lysed in lysis solution (10 mM Tris-HCl, pH 7.4, 3 mM MgCl₂, 10 mM NaCl, 150 mM sucrose, and 0.5% NP-40). Nuclei were isolated and suspended in 150 μ l of storage buffer (50 mM Tris-HCl, pH 8.3, 40% glycerol, 5 mM MgCl₂, and 0.1 mM EDTA) and 150 μ l 2x transcription buffer (300 mM KCl, 10 mM Tris-HCl, pH 8, 5 mM MgCl₂, 1 mM DTT, 500 μ M ATP, 500 μ M GTP, 500 μ M Br-UTP, 2 μ M CTP, and 200 U/ml Superase-in). After incubation at 30 °C for 30 min, 6 μ l stop buffer and 60 U RNase-free DNase I were added. RNAs were isolated using TRIzol. In order to isolate the newly synthesized RNA, Protein G beads were added, and incubated for 2 h at 37 °C. Beads were washed and RNAs were extracted

using TRIzol. Newly synthesized mRNA levels were measured by quantitative real-time PCR analysis.

Chromatin immunoprecipitation assay

Cells were fixed with 1% formaldehyde (Sigma-Aldrich) for 15 min at room temperature with shaking. Glycine (Sigma-Aldrich) was then added at a final concentration of 125 mM. The crosslinked cells were collected and prepared using lysis buffer (50 mM Tris, pH 8.1, 10 mM EDTA, 1% SDS, Roche complete protease inhibitor (EDTA-free)) with sonication. Equal amounts of protein were used for each immunoprecipitation reaction. Antibodies against histone H3 (ab1791), the RNA Pol II C-terminal domain (phospho-S2; ab5095), the RNA Pol II C-terminal domain (phospho-S5; ab5131), and histone H3 (tri-methyl K4, ab8580) were purchased from Abcam. Antibody against histone H3 acetyl Lys9 (39917) was purchased from Activemotif. Antibody against p300 (sc-48343) was purchased from Santa Cruz Biotechnology. The chromatin immunoprecipitation reaction was carried out with 2 μ L of antibody. Immunoprecipitated DNA was enriched using GammaBind G Sepharose beads (GE Healthcare) and eluted using elution buffer. Purified DNA was quantified by real-time qPCR. Occupancies were normalized by the ratio of ChIP to Input DNA.

In vitro transcription

To prepare the templates for in vitro transcription, the plasmids were linearized by NheI followed by successive phenol-chloroform extraction and ethanol precipitation. The capped and poly-A tailed mRNA transcripts were synthesized using HiScribe T7 quick high yield RNA

synthesis kit (NEB) supplemented with 3'-o-Me-m⁷G(5')ppp(5')G anti-reverse cap structure analog (NEB) following manufacturer's instructions. The mRNA concentrations were measured using a Nanodrop (Thermo Scientific).

In vitro translation using mammalian HEK-293T and *S. cerevisiae* cell-free lysates

To prepare *HEK-293T* cell-free lysate, HEK-293T cells were harvested by centrifugation at $1000 \times g$ for 4 min and washed with PBS three times. Cell pellets were resuspended in 2 volumes of hypotonic buffer (10 mM HEPES-KOH, pH 7.6, 10 mM potassium acetate, 0.5 mM magnesium acetate, 5 mM dithiothreitol), and incubated on ice for 40 min to 1 h. Cells were then homogenized by 20–30 strokes in a Dounce homogenizer on ice, and the final concentration of potassium acetate was adjusted to 50 mM. The cell extract was centrifuged at $16,000 \times g$ for 10 min at 4 °C. The supernatant was aliquoted, snap frozen in liquid nitrogen, and stored at –80 °C before use. To perform translation assay, 3 μ l reaction mixture (20 mM HEPES-KOH pH 7.6, 0.5 mM spermidine, 8 mM creatine phosphate, 0.2 mM GTP, 1 mM ATP, 20 μ M complete amino acids [Promega], 100 mM potassium acetate, 1 mM magnesium acetate, 0.13U/ μ l creatine phosphate kinase, 0.2U/ μ l SUPERaseIn RNase Inhibitor [Invitrogen]), 1 μ l of the mRNA template (180 ng) and 8 μ l cell-free translation extract I was used in each reaction. The reactions were incubated in 30°C water bath for 30 min and stopped by adding sodium dodecyl sulphate (SDS) sample buffer, followed immediately by heating at 90°C. The samples were subsequently analyzed by western blot.

To prepare *S. cerevisiae* cell free lysate, cells were harvested by centrifugation at 4° for 5 min at 3000 rpm, and resuspended in 1.5 ml of buffer A (30 mM HEPES-KOH, pH 7.6,

100 mM potassium acetate, 3 mM magnesium acetate; 2 mM dithiothreitol) with 8.5% mannitol and 0.5 mM PMSF per gram of cell weight (Wu et al. 2007). Lysate were centrifuged at 4° for 6 min at 18,000 rpm and supernatant was collected. Small molecular weight molecules are removed from the extract using Zeba Desalt Spin Columns (Pierce). Aliquots (200 µl) are pipetted into 1.6-ml Eppendorf tubes, frozen with liquid nitrogen, and stored at -80°. To perform translation assay, 7 µl translation reaction mixture [5 µl of cell lysate with 1 µl of 10x energy mix, 0.06 µl of 10U creatine phosphate kinase, 0.5 µl of 2M KOAc, 0.12 µl of 0.1M Mg(OAc)₂, 0.1 µl of 1 mM amino acids mix and 0.1 µl of SUPERase.In RNase inhibitor (Life Technologies), and 0.12 µl of RNase-free water], and 3 µl of the mRNA template (60 ng) was used in each reaction. The reactions were incubated in 26°C water bath for 15 min and stopped by adding SDS sample buffer, followed immediately by heating at 90°C. The samples were subsequently analyzed by western blot.

Polysome profiling

HEK-293T cells stably expressing Kras and Kras* were suspended in 425 µl of hypotonic buffer (5 mM Tris-HCL, pH 7.5, 2.5 mM MgCl₂, 1.5 mM KCl, Roche complete protease inhibitor (EDTA-free)) and 5 µl of 10 mg/ml CHX, 1 µl of 1 M DTT, 100 units of RNasin were added, and the samples were vortexed. After 5 min, 25 µl of 10% Triton X-100 and 25 µl of 10% sodium deoxycholate were added, and samples were centrifuged at 13,000 g for 10 min at 4 °C. The collected supernatants were then loaded onto a sucrose gradient prepared in 200 mM HEPES, pH 7.6, 1 M KCl, 50 mM MgCl₂, 100 µg/ml CHX, Roche complete protease inhibitor (EDTA-free), and 100 units of RNasin and centrifuged at 35,000 g for 2 h at 4 °C.

Fractions were collected, and absorbance at 254 nm was monitored to obtain the polysome profiles. RNA samples were isolated from individual fractions using the TRIzol reagent (Invitrogen) and resolved by northern blot.

BIBLIOGRAPHY

- Abruzzi KC, Rodriguez J, Menet JS, Desrochers J, Zadina A, Luo W, Tkachev S, Rosbash M. 2011. Drosophila CLOCK target gene characterization: implications for circadian tissue-specific gene expression. *Genes & development* **25**: 2374-2386.
- Akashi H. 1994. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics* **136**: 927-935.
- Ali M, Kaltenbrun E, Anderson GR, Stephens SJ, Arena S, Bardelli A, Counter CM, Wood KC. 2017. Codon bias imposes a targetable limitation on KRAS-driven therapeutic resistance. *Nat Commun* **8**: 15617.
- Antonellis A, Green ED. 2008. The role of aminoacyl-tRNA synthetases in genetic diseases. *Annual review of genomics and human genetics* **9**: 87-107.
- Bae K, Lee C, Hardin PE, Edery I. 2000. dCLOCK is present in limiting amounts and likely mediates daily interactions between the dCLOCK-CYC transcription factor and the PER-TIM complex. *J Neurosci* **20**: 1746-1753.
- Bali V, Bebok Z. 2015. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *Int J Biochem Cell B* **64**: 58-74.
- Bartoszewski RA, Jablonsky M, Bartoszewska S, Stevenson L, Dai Q, Kappes J, Collawn JF, Bebok Z. 2010. A Synonymous Single Nucleotide Polymorphism in Delta F508 CFTR Alters the Secondary Structure of the mRNA and the Expression of the Mutant Protein. *Journal of Biological Chemistry* **285**: 28741-28748.
- Bazzini AA, Del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, Yao J, Khokha MK, Giraldez AJ. 2016. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J*.
- Bennetzen JL, Hall BD. 1982. Codon selection in yeast. *The Journal of biological chemistry* **257**: 3026-3031.
- Birkeland E, Wik E, Mjos S, Hoivik EA, Trovik J, Werner HM, Kusonmano K, Petersen K, Raeder MB, Holst F et al. 2012. KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer. *Br J Cancer* **107**: 1997-2004.
- Blau J. 2008. PERSpective on PER phosphorylation. *Genes & development* **22**: 1737-1740.
- Boel G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff JD, Valecha M, Everett JK, Acton TB et al. 2016. Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature* **529**: 358-363.
- Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, Rodnina MV, Komar AA. 2016. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol Cell* **61**: 341-351.
- Carpen JD, von Schantz M, Smits M, Skene DJ, Archer SN. 2006. A silent polymorphism in the PER1 gene associates with extreme diurnal preference in humans. *Journal of human genetics* **51**: 1122-1125.
- Chang DC, Reppert SM. 2003. A novel C-terminal domain of drosophila PERIOD inhibits dCLOCK:CYCLE-mediated transcription. *Curr Biol* **13**: 758-762.

- Chiu JC, Ko HW, Edery I. 2011. NEMO/NLK phosphorylates PERIOD to initiate a time-delay phosphorylation circuit that sets circadian clock speed. *Cell* **145**: 357-370.
- Chiu JC, Low KH, Pike DH, Yildirim E, Edery I. 2010. Assaying locomotor activity to study circadian rhythms and sleep parameters in *Drosophila*. *J Vis Exp*.
- Chiu JC, Vanselow JT, Kramer A, Edery I. 2008. The phospho-occupancy of an atypical SLIMB-binding site on PERIOD that is phosphorylated by DOUBLETIME controls the pace of the clock. *Genes & development* **22**: 1758-1772.
- Cusumano P, Klarsfeld A, Chelot E, Picot M, Richier B, Rouyer F. 2009. PDF-modulated visual inputs and cryptochrome define diurnal behavior in *Drosophila*. *Nature neuroscience* **12**: 1431-1437.
- Cyran SA, Yiannoulos G, Buchsbaum AM, Saez L, Young MW, Blau J. 2005. The double-time protein kinase regulates the subcellular localization of the *Drosophila* clock protein period. *J Neurosci* **25**: 5430-5437.
- Dittmar KA, Goodenbour JM, Pan T. 2006. Tissue-specific differences in human transfer RNA expression. *PLoS Genet* **2**: e221.
- Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**: 341-352.
- Dunker AK, Silman I, Uversky VN, Sussman JL. 2008. Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* **18**: 756-764.
- Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**: 197-208.
- Fu J, Murphy KA, Zhou M, Li YH, Lam VH, Tabuloc CA, Chiu JC, Liu Y. 2016. Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev* **30**: 1761-1775.
- Garbe DS, Fang Y, Zheng X, Sowcik M, Anjum R, Gygi SP, Sehgal A. 2013. Cooperative interaction between phosphorylation sites on PERIOD maintains circadian period in *Drosophila*. *PLoS genetics* **9**: e1003749.
- Gingold H, Pilpel Y. 2011. Determinants of translation efficiency and accuracy. *Mol Syst Biol* **7**: 481.
- Hennig S, Strauss HM, Vanselow K, Yildiz O, Schulze S, Arens J, Kramer A, Wolf E. 2009. Structural and functional analyses of PAS domain interactions of the clock proteins *Drosophila* PERIOD and mouse PERIOD2. *PLoS biology* **7**: e94.
- Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet* **42**: 287-299.
- Hsin JP, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**: 2119-2137.
- Huang ZJ, Edery I, Rosbash M. 1993. PAS is a dimerization domain common to *Drosophila* period and several transcription factors. *Nature* **364**: 259-262.
- Iwatsubo T. 2000. Protein aggregation in neurodegenerative disorders: Alzheimer's disease and Parkinson's disease. *Int Congr Ser*: 137-138.
- Jafari R, Almqvist H, Axelsson H, Ignatushchenko M, Lundback T, Nordlund P, Molina DM. 2014. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat Protoc* **9**: 2100-2122.

- Kadener S, Stoleru D, McDonald M, Nawatheat P, Rosbash M. 2007. Clockwork Orange is a transcriptional repressor and a new *Drosophila* circadian pacemaker component. *Genes & development* **21**: 1675-1686.
- Kim EY, Ko HW, Yu WJ, Hardin PE, Edery I. 2007. A DOUBLETIME kinase binding domain on the *Drosophila* PERIOD protein is essential for its hyperphosphorylation, transcriptional repression, and circadian clock function. *Mol Cell Biol* **27**: 5014-5028.
- Kim SJ, Yoon JS, Shishido H, Yang Z, Rooney LA, Barral JM, Skach WR. 2015. Translational tuning optimizes nascent protein folding in cells. *Science* **348**: 444-448.
- Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**: 525-528.
- Kirchner S, Cai Z, Rauscher R, Kastelic N, Anding M, Czech A, Kleizen B, Ostedgaard LS, Braakman I, Sheppard DN et al. 2017. Alteration of protein function by a silent polymorphism linked to tRNA abundance. *PLoS Biol* **15**: e2000779.
- Kivimae S, Saez L, Young MW. 2008. Activating PER repressor through a DBT-directed phosphorylation switch. *PLoS Biol* **6**: e183.
- Ko HW, DiMassa S, Kim EY, Bae K, Edery I. 2007. Cis-combination of the classic per(S) and per(L) mutations results in arrhythmic *Drosophila* with ectopic accumulation of hyperphosphorylated PERIOD protein. *J Biol Rhythms* **22**: 488-501.
- Ko HW, Jiang J, Edery I. 2002. Role for Slimb in the degradation of *Drosophila* Period protein phosphorylated by Doubletime. *Nature* **420**: 673-678.
- Ko HW, Kim EY, Chiu J, Vanselow JT, Kramer A, Edery I. 2010. A hierarchical phosphorylation cascade that regulates the timing of PERIOD nuclear entry reveals novel roles for proline-directed kinases and GSK-3 β /SGG in circadian clocks. *J Neurosci* **30**: 12664-12675.
- Komar AA, Lesnik T, Reiss C. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett* **462**: 387-391.
- Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452-2460.
- Krinner S, Heitzer AP, Diermeier SD, Obermeier I, Langst G, Wagner R. 2014. CpG domains downstream of TSSs promote high levels of gene expression. *Nucleic Acids Res* **42**: 3551-3564.
- Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol* **4**: e180.
- Lampson BL, Pershing NL, Prinz JA, Lacsina JR, Marzluff WF, Nicchitta CV, MacAlpine DM, Counter CM. 2013. Rare codons regulate KRas oncogenesis. *Curr Biol* **23**: 70-75.
- Lazrak A, Fu L, Bali V, Bartoszewski R, Rab A, Havasi V, Keiles S, Kappes J, Kumar R, Lefkowitz E et al. 2013a. The silent codon change I507-ATC->ATT contributes to the severity of the DeltaF508 CFTR channel dysfunction. *FASEB J* **27**: 4630-4645.
- Lazrak A, Fu LW, Bali V, Bartoszewski R, Rab A, Havasi V, Keiles S, Kappes J, Kumar R, Lefkowitz E et al. 2013b. The silent codon change I507-ATC -> ATT contributes to the severity of the Delta F508 CFTR channel dysfunction. *Faseb J* **27**: 4630-4645.

- Lear BC, Zhang L, Allada R. 2009. The neuropeptide PDF acts directly on evening pacemaker neurons to regulate multiple features of circadian behavior. *PLoS biology* **7**: e1000154.
- Lee C, Bae K, Edery I. 1998. The Drosophila CLOCK protein undergoes daily rhythms in abundance, phosphorylation, and interactions with the PER-TIM complex. *Neuron* **21**: 857-867.
- . 1999. PER and TIM inhibit the DNA binding activity of a Drosophila CLOCK-CYC/DBMAL1 heterodimer without disrupting formation of the heterodimer: a basis for circadian transcription. *Mol Cell Biol* **19**: 5316-5325.
- Lee JW, Beebe K, Nangle LA, Jang JS, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P, Ackerman SL. 2006. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443**: 50-55.
- Li Y, Guo F, Shen J, Rosbash M. 2014. PDF and cAMP enhance PER stability in Drosophila clock neurons (vol 111, pg E1284, 2014). *Proceedings of the National Academy of Sciences of the United States of America* **111**: 8311-8311.
- Liu ML, Zang T, Zhang CL. 2016. Direct Lineage Reprogramming Reveals Disease-Specific Phenotypes of Motor Neurons from Human ALS Patients. *Cell reports* **14**: 115-128.
- Matsuo M, Shino Y, Yamada N, Ozeki Y, Okawa M. 2007. A novel SNP in hPer2 associates with diurnal preference in a healthy population. *Sleep Biol Rhythms* **5**: 141-145.
- McCarthy C, Carrea A, Diambra L. 2017. Bicondon bias can determine the role of synonymous SNPs in human diseases. *BMC Genomics* **18**: 227.
- Menet JS, Abruzzi KC, Desrochers J, Rodriguez J, Rosbash M. 2010. Dynamic PER repression mechanisms in the Drosophila circadian clock: from on-DNA to off-DNA. *Genes Dev* **24**: 358-367.
- Merbitz-Zahradnik T, Wolf E. 2015. How is the inner circadian clock controlled by interactive clock proteins?: Structural analysis of clock proteins elucidates their physiological role. *FEBS letters* **589**: 1516-1529.
- Mishima Y, Tomari Y. 2016. Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Mol Cell* **61**: 874-885.
- Molina DM, Jafari R, Ignatushchenko M, Seki T, Larsson EA, Dan C, Sreekumar L, Cao YH, Nordlund P. 2013. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **341**: 84-87.
- Muskus MJ, Preuss F, Fan JY, Bjes ES, Price JL. 2007. Drosophila DBT lacking protein kinase activity produces long-period and arrhythmic circadian behavioral and molecular rhythms. *Mol Cell Biol* **27**: 8049-8064.
- Nagoshi E, Sugino K, Kula E, Okazaki E, Tachibana T, Nelson S, Rosbash M. 2010. Dissecting differential gene expression within the circadian neuronal circuit of Drosophila. *Nature neuroscience* **13**: 60-68.
- Newman ZR, Young JM, Ingolia NT, Barton GM. 2016. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc Natl Acad Sci U S A* **113**: E1362-1371.
- Pechmann S, Chartron JW, Frydman J. 2014. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol* **21**: 1100-1105.

- Pechmann S, Frydman J. 2013. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* **20**: 237-243.
- Pershing NL, Lampson BL, Belsky JA, Kaltenbrun E, MacAlpine DM, Counter CM. 2015. Rare codons capacitate Kras-driven de novo tumorigenesis. *J Clin Invest* **125**: 222-233.
- Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12**: 32-42.
- Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR et al. 2015. Codon optimality is a major determinant of mRNA stability. *Cell* **160**: 1111-1124.
- Preuss F, Fan JY, Kalive M, Bao S, Schuenemann E, Bjes ES, Price JL. 2004. Drosophila doubletime mutations which either shorten or lengthen the period of circadian rhythms decrease the protein kinase activity of casein kinase I. *Mol Cell Biol* **24**: 886-898.
- Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. 2012. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet* **8**: e1002603.
- Radhakrishnan A, Chen YH, Martin S, Alhusaini N, Green R, Collier J. 2016. The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell* **167**: 122-132 e129.
- Rosato E, Kyriacou CP. 2006. Analysis of locomotor activity rhythms in Drosophila. *Nat Protoc* **1**: 559-568.
- Ross CA, Poirier MA. 2004. Protein aggregation and neurodegenerative disease. *Nat Med* **10**: S10-S17.
- Sander IM, Chaney JL, Clark PL. 2014. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc* **136**: 858-861.
- Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**: 683-691.
- Schmitt BM, Rudolph KLM, Karagianni P, Fonseca NA, White RJ, Talianidis L, Odom DT, Marioni JC, Kutter C. 2014. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome Res* **24**: 1797-1807.
- Sharp PM, Li W-H. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15**: 1281-1295.
- Siller E, DeZwaan DC, Anderson JF, Freeman BC, Barral JM. 2010. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J Mol Biol* **396**: 1310-1318.
- Spencer PS, Siller E, Anderson JF, Barral JM. 2012. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol* **422**: 328-335.
- Stoleru D, Peng Y, Agosto J, Rosbash M. 2004. Coupled oscillators control morning and evening locomotor behaviour of Drosophila. *Nature* **431**: 862-868.
- Taylor P, Hardin PE. 2008. Rhythmic E-box binding by CLK-CYC controls daily cycles in *per* and *tim* transcription and chromatin modifications. *Mol Cell Biol* **28**: 4642-4652.
- Tomba P. 2011. Unstructural biology coming of age. *Curr Opin Struct Biol* **21**: 419-425.
- Waldman YY, Tuller T, Shlomi T, Sharan R, Ruppin E. 2010. Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages. *Nucleic acids research* **38**: 2964-2974.

- Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP. 2016. Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell reports* **14**: 1787-1799.
- Wu C, Amrani N, Jacobson A, Sachs MS. 2007. The use of fungal in vitro systems for studying translational regulation. *Methods Enzymol* **429**: 203-225.
- Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, Liu Y. 2015. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol Cell* **59**: 744-754.
- Yu Q, Jacquier AC, Citri Y, Hamblen M, Hall JC, Rosbash M. 1987. Molecular mapping of point mutations in the period gene that stop or speed up biological clocks in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **84**: 784-788.
- Zhang G, Hubalewska M, Ignatova Z. 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* **16**: 274-280.
- Zhao F, Yu CH, Liu Y. 2017. Codon usage regulates protein structure and function by affecting translation elongation speed in *Drosophila* cells. *Nucleic Acids Res* **45**: 8484-8492.
- Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y. 2013. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**: 111-115.
- Zhou M, Wang T, Fu J, Xiao G, Liu Y. 2015. Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol Microbiol* **97**: 974-987.
- Zhou T, Weems M, Wilke CO. 2009. Translationally optimal codons associate with structurally sensitive sites in proteins. *Molecular biology and evolution* **26**: 1571-1580.
- Zhou Z, Dang Y, Zhou M, Li L, Yu CH, Fu J, Chen S, Liu Y. 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci U S A* **113**: E6117-E6125.