

NIH grant proposal

Douglas Myers-Turnbull, Robin Betz, Yunsup Jung

Things to do

1: Disturbingly, there's a bug in the TODOs. Why are they all numbered with '1'?

Contents

Specific aims	1
Background	1
Proposal	2
Aim 1: Develop a comprehensive, verifiable, and rigorous pipeline to elucidate relationships between codon usage bias and protein structure.	3
Aim 2: Identify clinically significant effects of synonymous mutations on protein structure.	3
Aim 3: Establish codon usage bias as an important biological mechanism.	3
Research strategy	3
Significance	3
Innovation	3
Approach	3
Preliminary data	3
Pipeline development	3
Assessment of accuracy	3
Clinical studies	3
Folding study	3

TODO: Disturbingly, there's a bug in the TODOs. Why are they all numbered with '1'?

Specific aims

Background

It is known that codon usage bias correlates with expression levels [7, 11]. There are strong indications that abundance of isoaccepting tRNA molecules [12, 16, 18, 20] is a causal factor of this differential expression. Specifically, proteins with high expression levels have been found to contain greater levels of commonly used

codons, and frequently used codons are associated with higher levels of corresponding tRNAs. Thus there is believed to be a positive selection for a small, constrained set of codon–tRNA combinations, and concentrations of codons and tRNAs are related in a positive feedback cycle, where bias in either causes a positive selection for bias in the other. Because of this strong correlation, we call infrequently used codons *slow*, and frequently used codons *fast*, except in cases where we address this correlation directly. Furthermore, we consider sequences containing a large proportion of frequently used codons to have high *codon usage bias*, and sequences containing either a moderate or low proportion to be *unbiased*, even though sequences with low proportion still deviate from the statistical mean.

There is also substantial evidence that translation dynamics is responsible for some of this correlation [2, 4, 5, 6, 14, 15]. Factors including mRNA secondary structure, mRNA stability [10], and codon–tRNA affinity have been suggested. Most notably, these investigations have shown that ribosomal traffic is an important factor affecting codon usage bias. For efficient translation, overall codon usage should be skewed in favor of frequent (fast) codons, and there should be a gradient in which slow codons are prevalent at the beginning of the transcript but rare high toward the middle and end. Mitarai and Pedersen [15] used a simple computational model to show that the introduction of even a single slow codon near the end of the transcript can cause ribosomal traffic jams that drastically decrease translation speed.

Because slow codons can cause pauses in translation, it has been suggested that such pauses can influence protein folding, leading to different folded states for transcripts with differing synonymous codon usage [4, 13, 21]. Many studies have found that codon usage bias is correlated to protein structure [1, 3, 8, 19], including secondary structure [8, 17] and domain organization [9, 17]. More surprising is the recent demonstration by Zhou et al. [23] that differential codon usage can directly alter the folded state of a protein product. Zhou et al. changed the folded state of the circadian rhythm protein FRQ by introducing synonymous mutations.

Proposal

In general, the bioinformatics studies by Adzhubei et al. [1], Biro [3], Gu et al. [8], Saunders and Deane [19] controlled for very few variables and were able to identify only a few clear correlations, particularly between codon usage bias and protein secondary structure. However, we hypothesize that these results were negative because the effects of codon usage bias on protein structure are minor or nonexistent for most proteins. However, as shown by Zhou et al. [23], differential codon usage can dramatically affect the folding of certain proteins, and some of these, including FRQ, may be clinically important. Therefore, we further hypothesize that general mechanisms exist even if few general correlations do, and that controlling for more variables will allow us to elucidate general mechanisms.

We propose a large-scale bioinformatics study to identify causal relationships in which synonymous codon usage affects protein structure. To address the negative results of the studies by Adzhubei et al. [1], Biro [3], Gu et al. [8], Saunders and Deane [19], we intend to address causal relationships rather than correlations by developing a mathematically and statistically rigorous framework, which we will use to address the problem of weak signals.

Aim 1: Develop a comprehensive, verifiable, and rigorous pipeline to elucidate relationships between codon usage bias and protein structure.

Aim 2: Identify clinically significant effects of synonymous mutations on protein structure.

Aim 3: Establish codon usage bias as an important biological mechanism.

Research strategy

Significance

Innovation

Approach

Preliminary data

Pipeline development

Assessment of accuracy

Clinical studies

Folding study

As an auxillary study, we will investigate the effects of synonymous mutations in detail for select cases using MD-based folding software. Although de novo folding is still largely unsolved, and de novo folding algorithms are still in their infancy, such an investigation may still reveal insight for some cases that is unavailable through other means. [22]

Bibliography & References Cited

- [1] a a Adzhubei, I a Adzhubei, I a Krasheninnikov, and S Neidle. Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. *FEBS letters*, 399(1-2):78–82, December 1996. ISSN 0014-5793. URL <http://www.ncbi.nlm.nih.gov/pubmed/8980124>.
- [2] Kajetan Bentele, Paul Saffert, Robert Rauscher, Zoya Ignatova, and Nils Blüthgen. Efficient translation initiation dictates codon usage at gene start. *Molecular systems biology*, 9(675):675, January 2013. ISSN 1744-4292. doi: 10.1038/msb.2013.32. URL <http://www.ncbi.nlm.nih.gov/pubmed/23774758>.
- [3] Jan Charles Biro. Indications that "codon boundaries" are physico-chemically defined and that protein-folding information is contained in the redundant exon bases. *Theoretical biology & medical modelling*, 3: 28, January 2006. ISSN 1742-4682. doi: 10.1186/1742-4682-3-28. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1560374&tool=pmcentrez&rendertype=abstract>.
- [4] J Ross Buchan and Ian Stansfield. Halting a cellular production line: responses to ribosomal pausing during translation. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 99(9):475–87, September 2007. ISSN 1768-322X. doi: 10.1042/BC20070037. URL <http://www.ncbi.nlm.nih.gov/pubmed/17696878>.

- [5] Gina Cannarozzi, Gina Cannarozzi, Nicol N Schraudolph, Mahamadou Faty, Peter von Rohr, Markus T Friberg, Alexander C Roth, Pedro Gonnet, Gaston Gonnet, and Yves Barral. A role for codon order in translation dynamics. *Cell*, 141(2):355–67, April 2010. ISSN 1097-4172. doi: 10.1016/j.cell.2010.02.036. URL <http://www.ncbi.nlm.nih.gov/pubmed/20403329>.
- [6] Kurt Fredrick and Michael Ibba. How the sequence of a gene can tune its translation. *Cell*, 141(2):227–9, April 2010. ISSN 1097-4172. doi: 10.1016/j.cell.2010.03.033. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2866089&tool=pmcentrez&rendertype=abstract>.
- [7] Regina M Goetz and Anders Fuglsang. Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from Escherichia coli. *Biochemical and biophysical research communications*, 327(1): 4–7, February 2005. ISSN 0006-291X. doi: 10.1016/j.bbrc.2004.11.134. URL <http://www.ncbi.nlm.nih.gov/pubmed/15629421>.
- [8] Wanjun Gu, Tong Zhou, Jianmin Ma, Xiao Sun, and Zuhong Lu. Folding type specific secondary structure propensities of synonymous codons. ..., *IEEE Transactions on*, 2(3):150–157, 2003. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1229599.
- [9] Wanjun Gu, Tong Zhou, Jianmin Ma, Xiao Sun, and Zuhong Lu. The relationship between synonymous codon usage and protein structure in Escherichia coli and Homo sapiens. *Bio Systems*, 73(2):89–97, March 2004. ISSN 0303-2647. doi: 10.1016/j.biosystems.2003.10.001. URL <http://www.ncbi.nlm.nih.gov/pubmed/15013221>.
- [10] Wanjun Gu, Tong Zhou, and Claus O Wilke. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS computational biology*, 6(2):e1000664, February 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000664. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2816680&tool=pmcentrez&rendertype=abstract>.
- [11] Claes Gustafsson, Sridhar Govindarajan, and Jeremy Minshull. Codon bias and heterologous protein expression. *Trends in biotechnology*, 22(7):346–53, July 2004. ISSN 0167-7799. doi: 10.1016/j.tibtech.2004.04.006. URL <http://www.ncbi.nlm.nih.gov/pubmed/15245907>.
- [12] Stefan Klumpp, Jiajia Dong, and Terence Hwa. On ribosome load, codon bias and protein abundance. *PloS one*, 7(11):e48542, January 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0048542. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3492488&tool=pmcentrez&rendertype=abstract>.
- [13] Anton a Komar. A pause for thought along the co-translational folding pathway. *Trends in biochemical sciences*, 34(1):16–24, January 2009. ISSN 0968-0004. doi: 10.1016/j.tibs.2008.10.002. URL <http://www.ncbi.nlm.nih.gov/pubmed/18996013>.
- [14] Monica Marin. Folding at the rhythm of the rare codon beat. *Biotechnology journal*, 3(8):1047–57, August 2008. ISSN 1860-7314. doi: 10.1002/biot.200800089. URL <http://www.ncbi.nlm.nih.gov/pubmed/18624343>.
- [15] Namiko Mitarai and Steen Pedersen. Control of ribosome traffic by position-dependent choice of synonymous codons. *Physical biology*, 10(5):056011, October 2013. ISSN 1478-3975. doi: 10.1088/1478-3975/10/5/056011. URL <http://www.ncbi.nlm.nih.gov/pubmed/24104350>.
- [16] Hamed Shateri Najafabadi, Jean Lehmann, and Mohammad Omid. Error minimization explains the codon usage of highly expressed genes in Escherichia coli. *Gene*, 387(1-2):150–5, January 2007. ISSN 0378-1119. doi: 10.1016/j.gene.2006.09.004. URL <http://www.ncbi.nlm.nih.gov/pubmed/17097242>.

- [17] Matej Oresic, Michael Dehn, Daniel Korenblum, and David Shalloway. Tracing specific synonymous codon-secondary structure correlations through evolution. *Journal of molecular evolution*, 56(4):473–84, April 2003. ISSN 0022-2844. doi: 10.1007/s00239-002-2418-x. URL <http://www.ncbi.nlm.nih.gov/pubmed/12664167>.
- [18] Joshua B Plotkin and Grzegorz Kudla. Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics*, 12(1):32–42, January 2011. ISSN 1471-0064. doi: 10.1038/nrg2899. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3074964&tool=pmcentrez&rendertype=abstract>.
- [19] Rhodri Saunders and Charlotte M Deane. Synonymous codon usage influences the local protein structure observed. *Nucleic acids research*, 38(19):6719–6728, 2010.
- [20] S Tavaré and B Song. Codon preference and primary sequence structure in protein-coding regions. *Bulletin of mathematical biology*, 51(1):95–115, January 1989. ISSN 0092-8240. URL <http://www.ncbi.nlm.nih.gov/pubmed/2706404>.
- [21] Gong Zhang, Magdalena Hubalewska, and Zoya Ignatova. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature structural & molecular biology*, 16(3):274–80, March 2009. ISSN 1545-9985. doi: 10.1038/nsmb.1554. URL <http://www.ncbi.nlm.nih.gov/pubmed/19198590>.
- [22] Yang Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9:40, 2008.
- [23] Mian Zhou, Jinhu Guo, Joonseok Cha, Michael Chae, She Chen, Jose M Barral, Matthew S Sachs, and Yi Liu. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, 495(7439):111–5, March 2013. ISSN 1476-4687. doi: 10.1038/nature11833. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3629845&tool=pmcentrez&rendertype=abstract>.