

Abstract Submitted  
for the MAR07 Meeting of  
The American Physical Society

**A Model of Codon Usage Bias** MORTEN KLOSTER, CHAO TANG,  
UCSF — The genetic code is degenerate; most amino acids can be encoded by from two to as many as six different codons. While one might expect these codons to be used with equal frequency, this turns out not to be the case—not only are some codons favored over others, but their usage can vary significantly between different genes in the same organism. Known causes of codon bias include differences in mutation rates as well as selection pressure related to the expression level of a gene, but the standard analysis methods can explain only a fraction of the observed codon usage variation. We here introduce an explicit model of codon usage bias, inspired by statistical physics. Combining this model with a maximum likelihood approach, we are able to clearly identify up to four different sources of bias in various genomes. We have applied the algorithm to *Saccharomyces cerevisiae* as well as 325 bacterial genomes, and in most cases our model explains essentially all observed variance.

Morten Kloster  
UCSF

Date submitted: 20 Nov 2006

Electronic form version 1.4