Optimizing Biochemical Pathways for Ethanol Production in Yeast

Wes Hatfield, Ph.D.

DIRECTOR
COMPUTATIONAL BIOLOGY RESEARCH LABORATORY
INSTITUTE FOR TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY
UNIVERSITY OF CALIFORNIA, IRVINE

SCIENTIFIC FOUNDER
CODA GENOMICS, INC.
LAGUNA HILLS, CA

An Interdisciplinary CODA/UC collaboration

Funded by a UC Discovery Grant
based on two UC patent technologies

CODA

and

Translation Engineering™

Computationally Optimized DNA Assembly

A new platform technology at the interface of the computational and life sciences that enables the self-assembly of synthetic genes

Translational Engineering

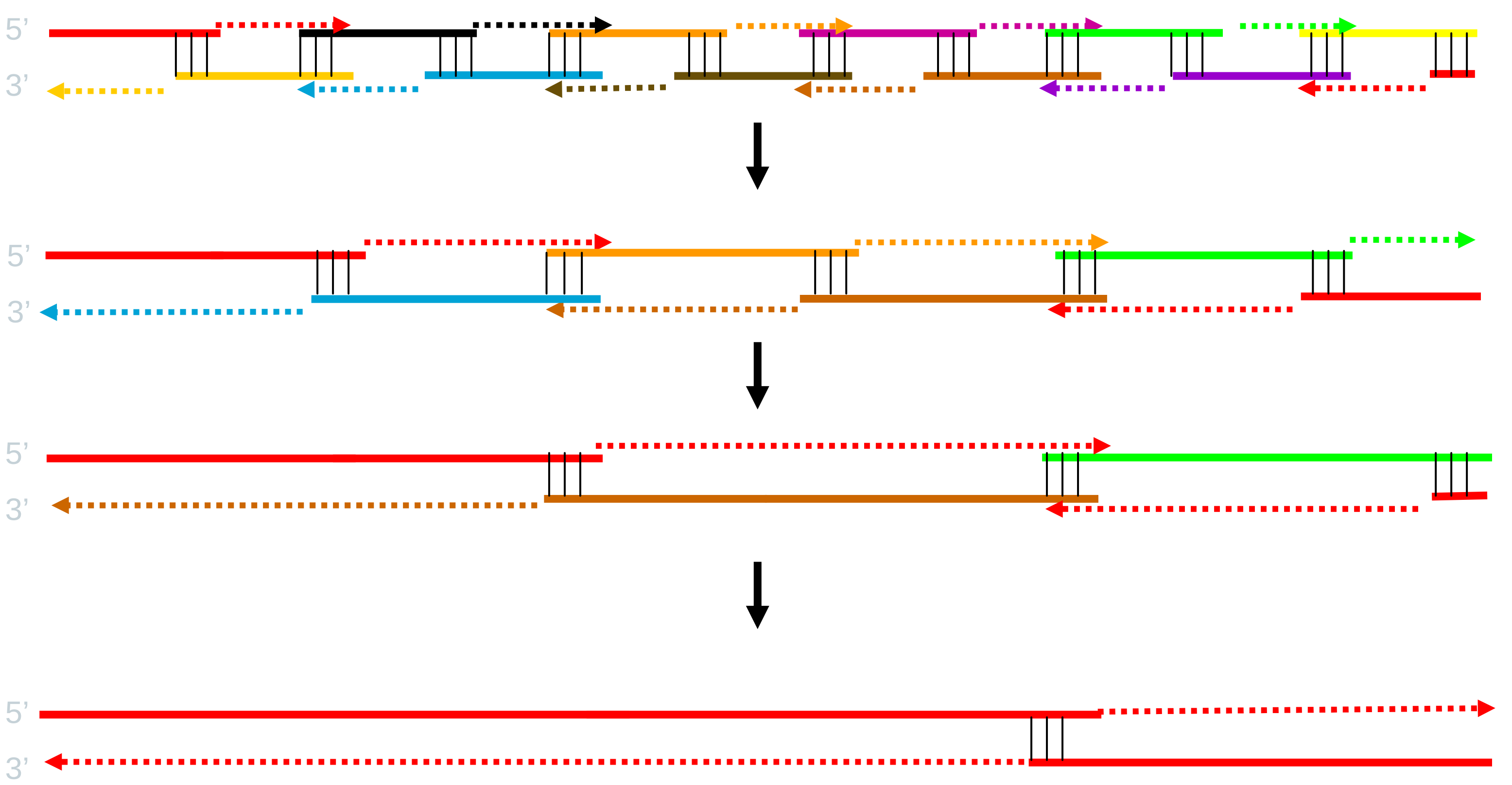
The use of codon pair utilization (codon context) patterns to affect translational step-times to improve protein expression levels and function in non-native organisms

Together these technologies:

- Enable rapid and accurate self-assembly of synthetic genes
- Enable the design of synthetic genes for optimal expression in any organism with a sequenced genome
- Enable the design of synthetic genes for the improved production of active proteins (enzymes) in any organism with a sequenced genome
- Licensed by UC to CODA Genomics, Inc., Laguna Hills, CA

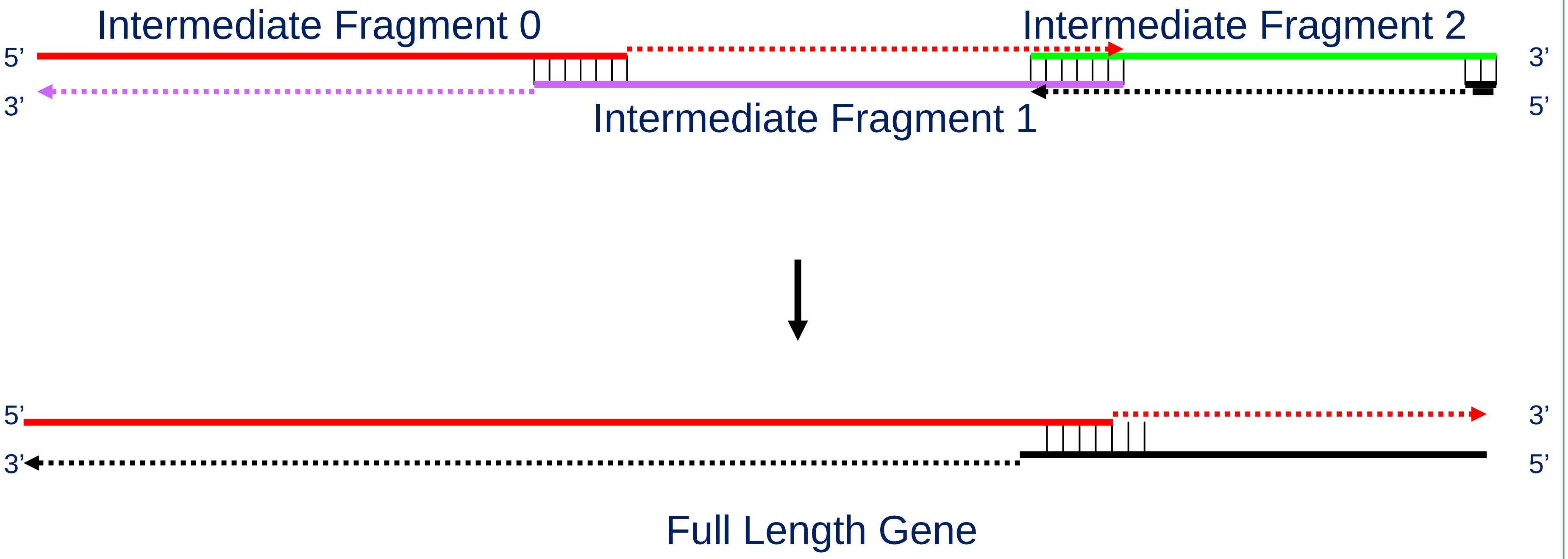
CODA

Synthetic genes are assembled from small oligos



Intermediate DNA Fragment

Full length genes are assembled from intermediate gene fragments



Problems:

- Incorrect oligos with point mutations and deletions are incorporated into the gene sequence
- oligos assemble in the wrong order

128 3GHz Xenon Processors

WHAT DOES CODA DO?

It optimizes the sequence of small, adjacent, overlapping oligos for self-assembly

ONLY correct sequence assembly

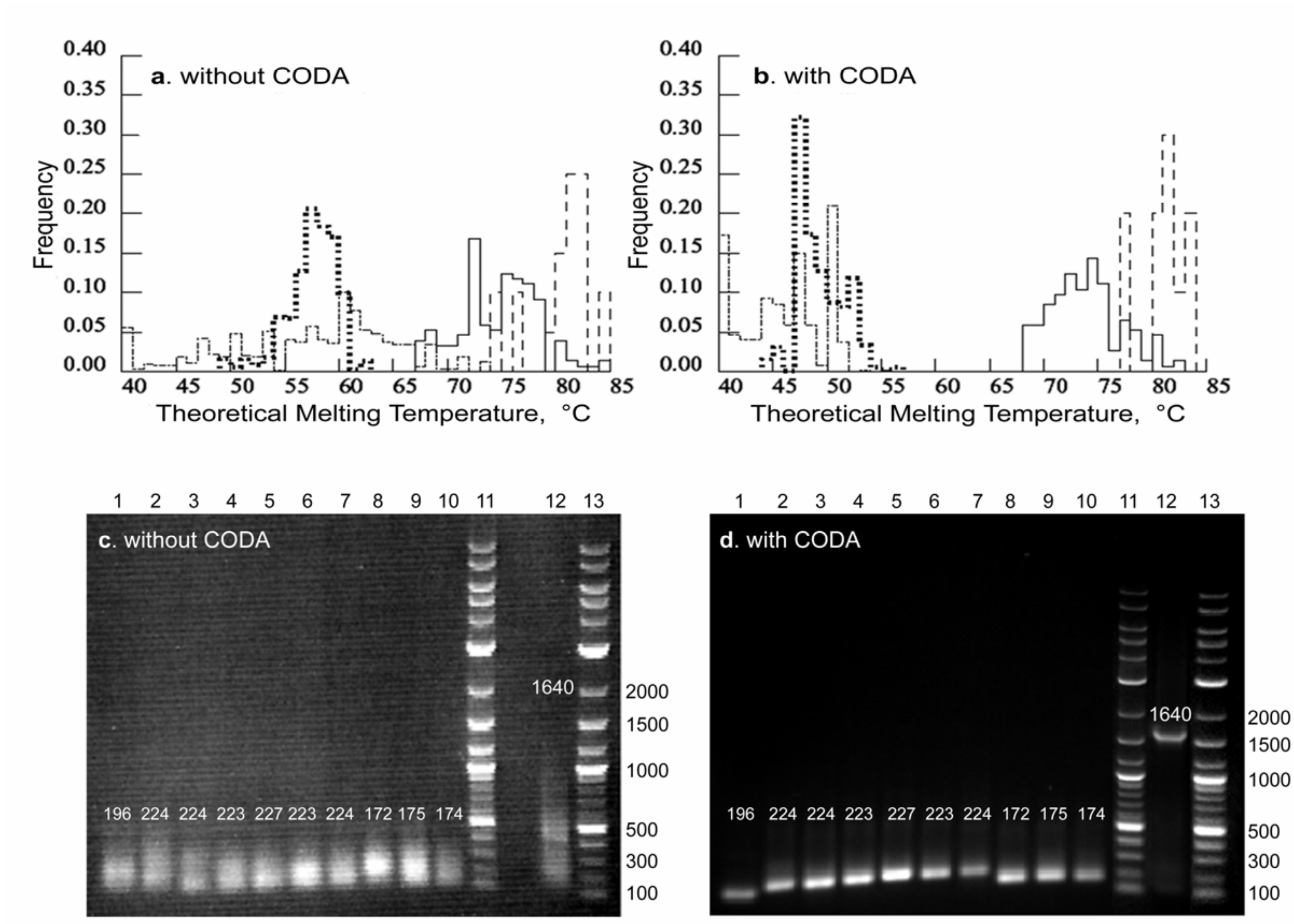
HOW DOES CODA DO IT?

It uses:

- The branch and bound (CODA) algorithm
  - » searches through large numbers of possible DNA sequences to find an admissible solution

"*Multi-Queue Branch-and-Bound Algorithm for Anytime Optimal Search with Biological Applications*," **Lathrop, R.H.**, Sazhin, A., Sun, Y., Steffen, N., Irani, S., (Best Paper Award), p. 73-82 in *Proc. Intl. Conf. on Genome Informatics,* Tokyo, Japan, Dec. 17-19, 2001.

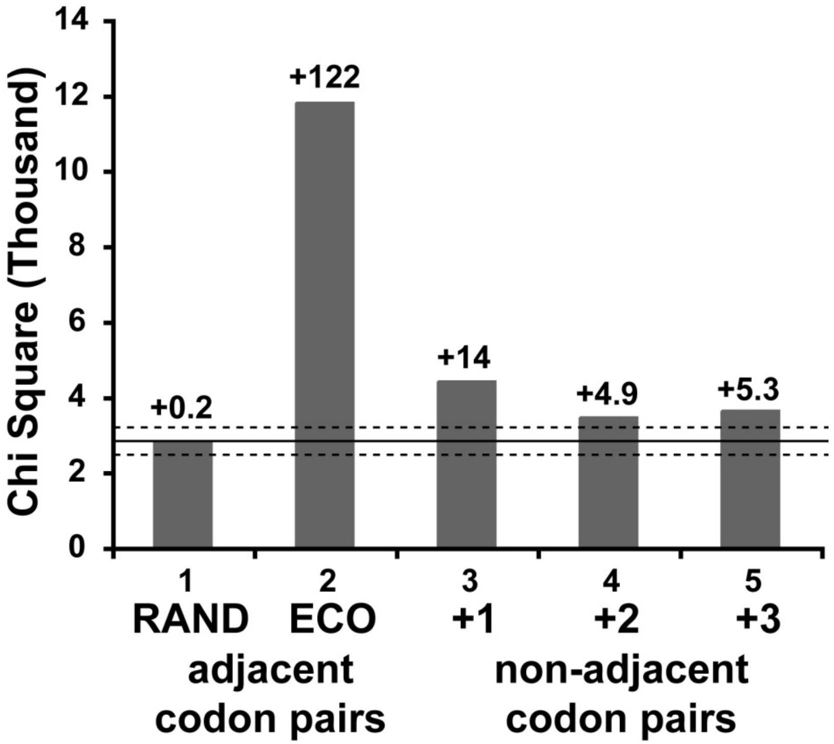
- Theoretical melting temperatures
  - » calculated for all possible correct and incorrect hybridizations
- Degeneracy of the genetic code
  - » strengthens correct hybridizations (high TM)
  - » disrupts incorrect hybridizations (low TM)
- Optimizes protein production in the host organism


a. without CODA
b. with CODA
Frequency
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
40 45 50 55 60 65 70 75 80 85
Theoretical Melting Temperature, °C
c. without CODA
d. with CODA
1 2 3 4 5 6 7 8 9 10 11 12 13
196 224 224 223 227 223 224 172 175 174
1640
2000
1500
1000
500
300
100

TRANSLATIONAL ENGINEERING

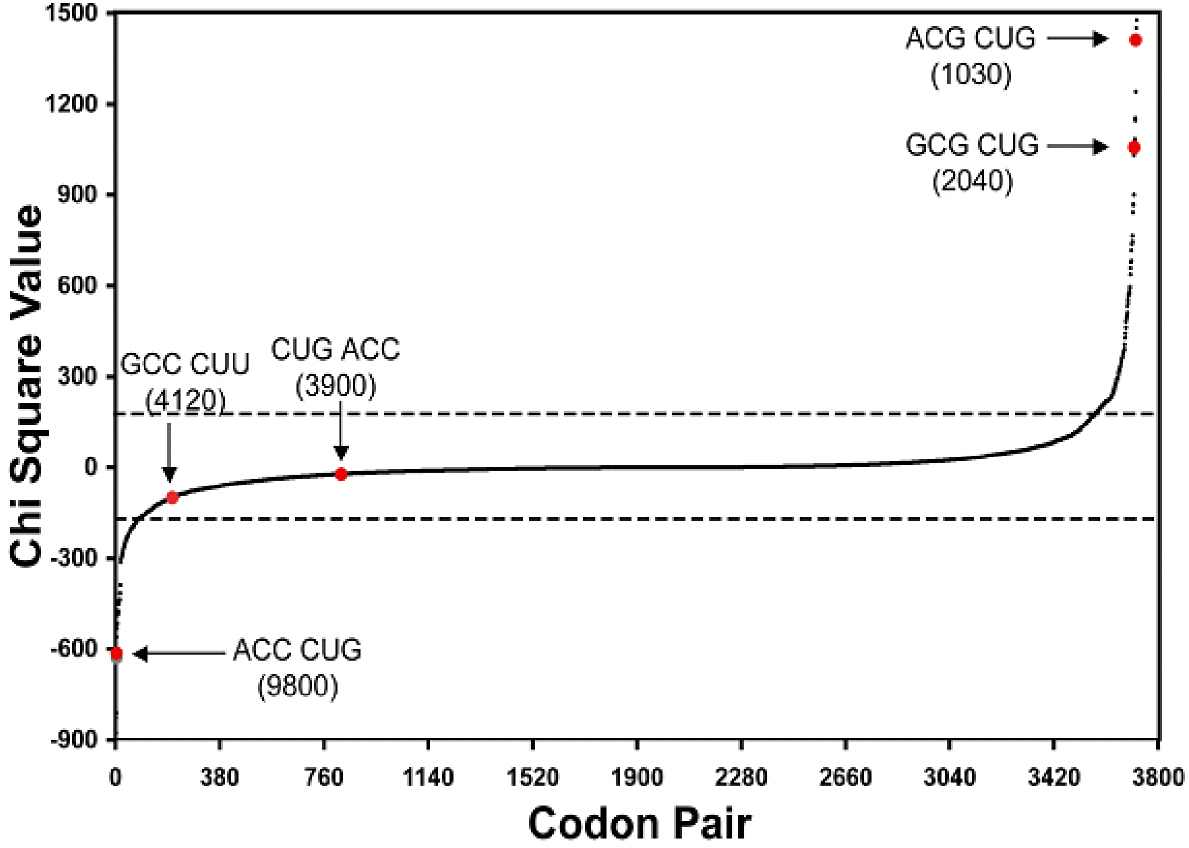
The use of codon pair utilization (codon context) patterns to affect translational step-times to improve protein expression levels and function

Codon Pair Bias (Codon Context) in *E. coli*



- Codon pair usage is highly biased in *E. coli* (and all organisms)
- The bias is limited to nearest neighbors.
- The bias is directional
- Codon pair bias is *unrelated* to codon usage

CODON PAIR BIAS

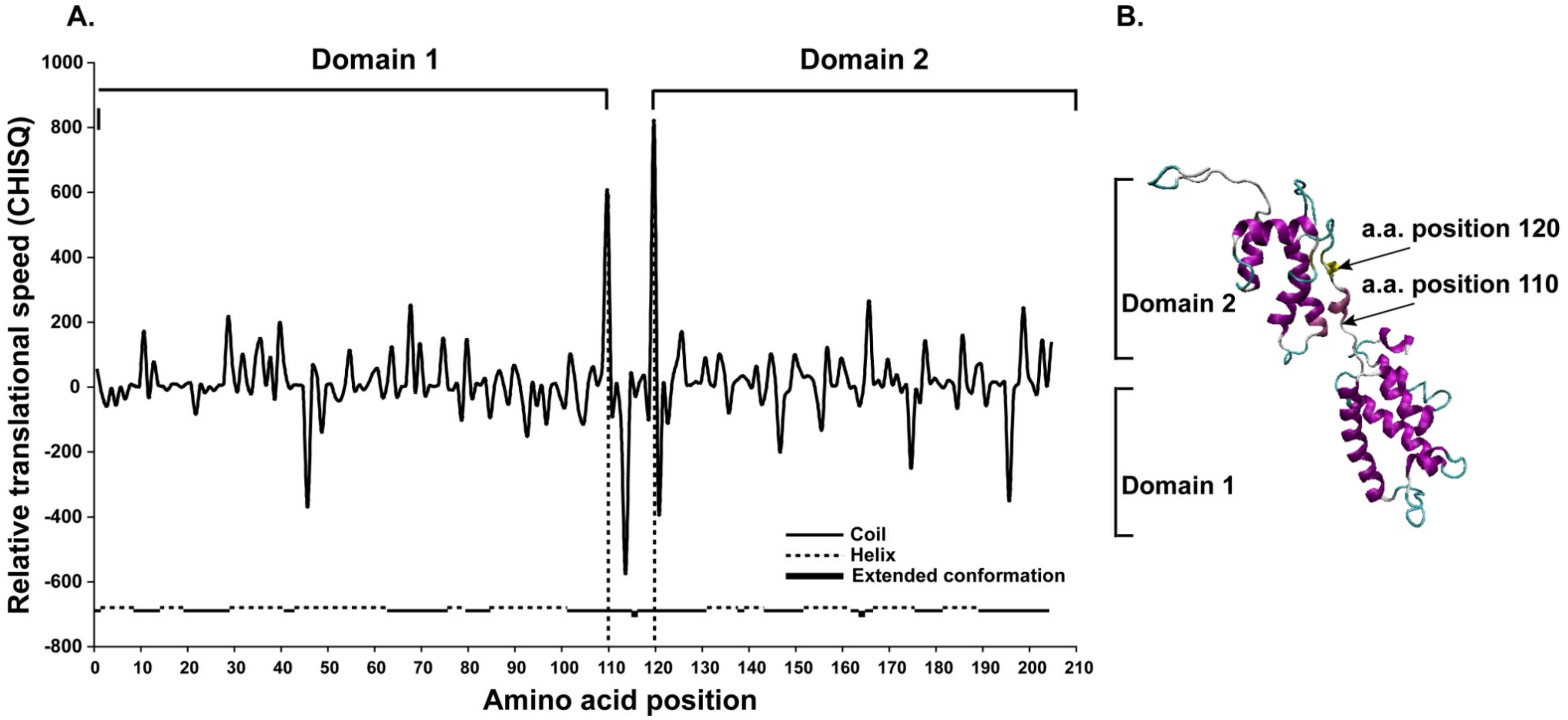


Irwin B, Heck JD, Hatfield GW. Codon pair utilization biases influence translational elongation step times. *J Biol Chem.* 1995 Sep 29;270(39):22801-6.

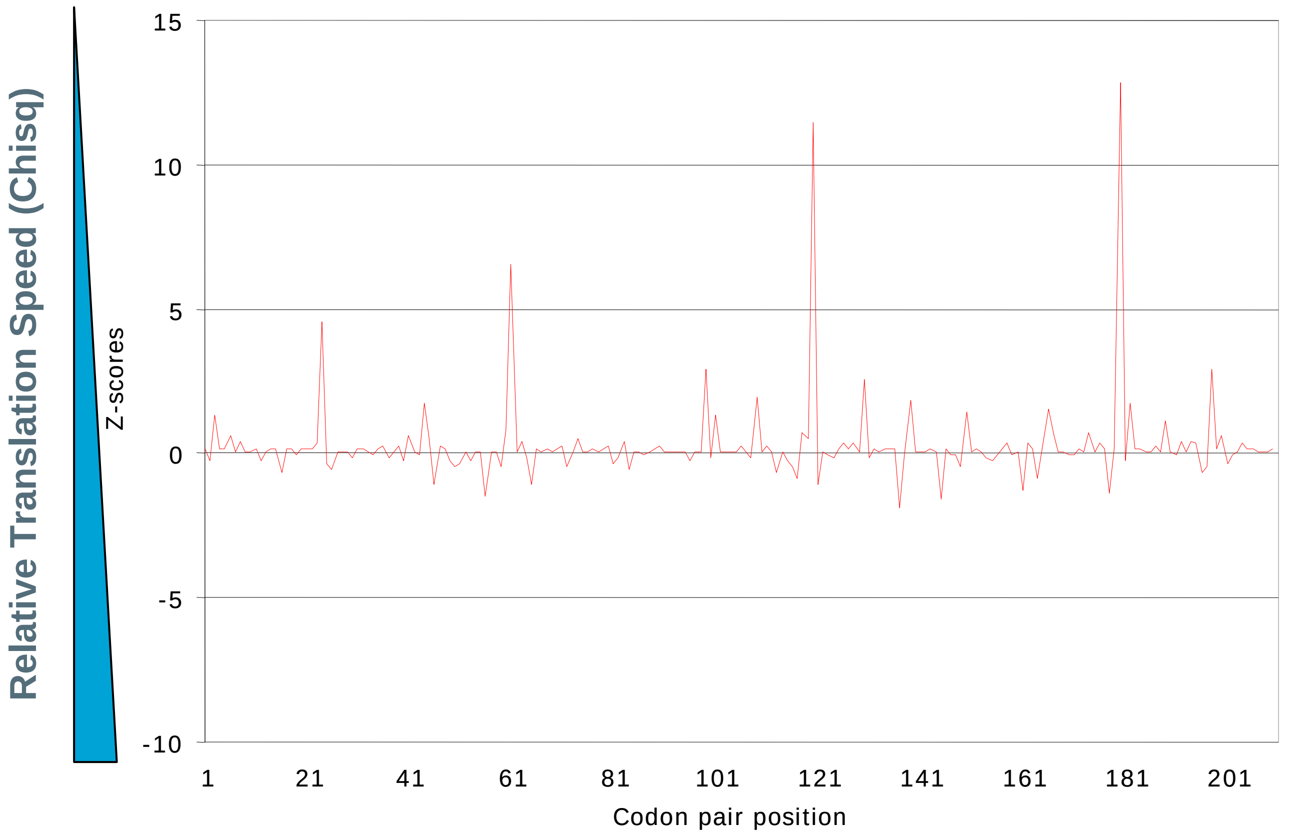
SUGGESTION:

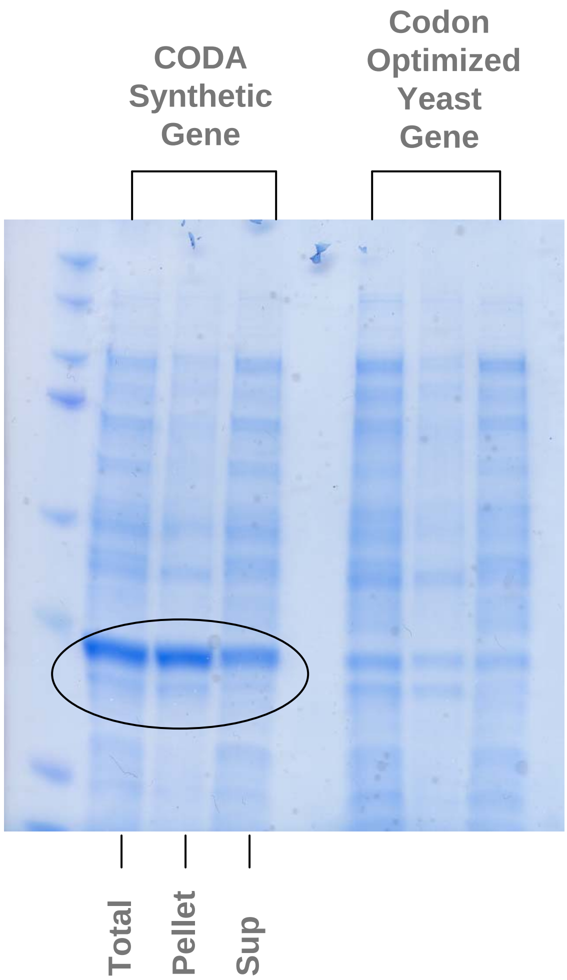
The frequency of one codon next to another has co-evolved with the structure and abundance of tRNA isoacceptors in order to control the rates of translational step times without imposing additional constraints on amino acid sequences or protein structures

N-terminal and C-terminal Capsid Domains of the *S. cerevisiae* Ty3 *GAG3* gene



SpeedPlot of Codon Optimized Yeast Ty3 *GAG3* gene in *E. coli*




CODA
Synthetic
Gene
Codon
Optimized
Yeast
Gene
Total
Pellet
Sup

CODA GENE SEQUENCES ARE OPTIMIZED FOR:

- Thermodynamics for self assembly
- Codon context properties for improved gene expression and protein function in non-native host organisms

Ethanol Production in Yeast
(*Saccharomyces cerevisiae*)

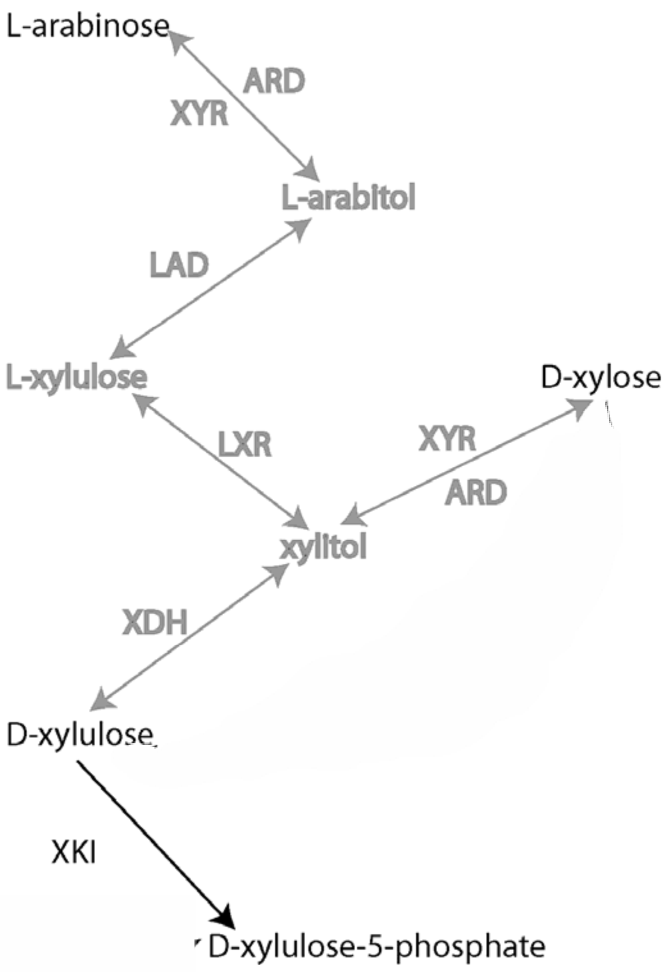
PROBLEMS:

- Yeast do not possess the biochemical pathways (enzymes) to use biomass (pentose) sugars
- Current engineered strains are inefficient:
  - Genes from other organisms do not work well
  - They suffer from redox imbalances

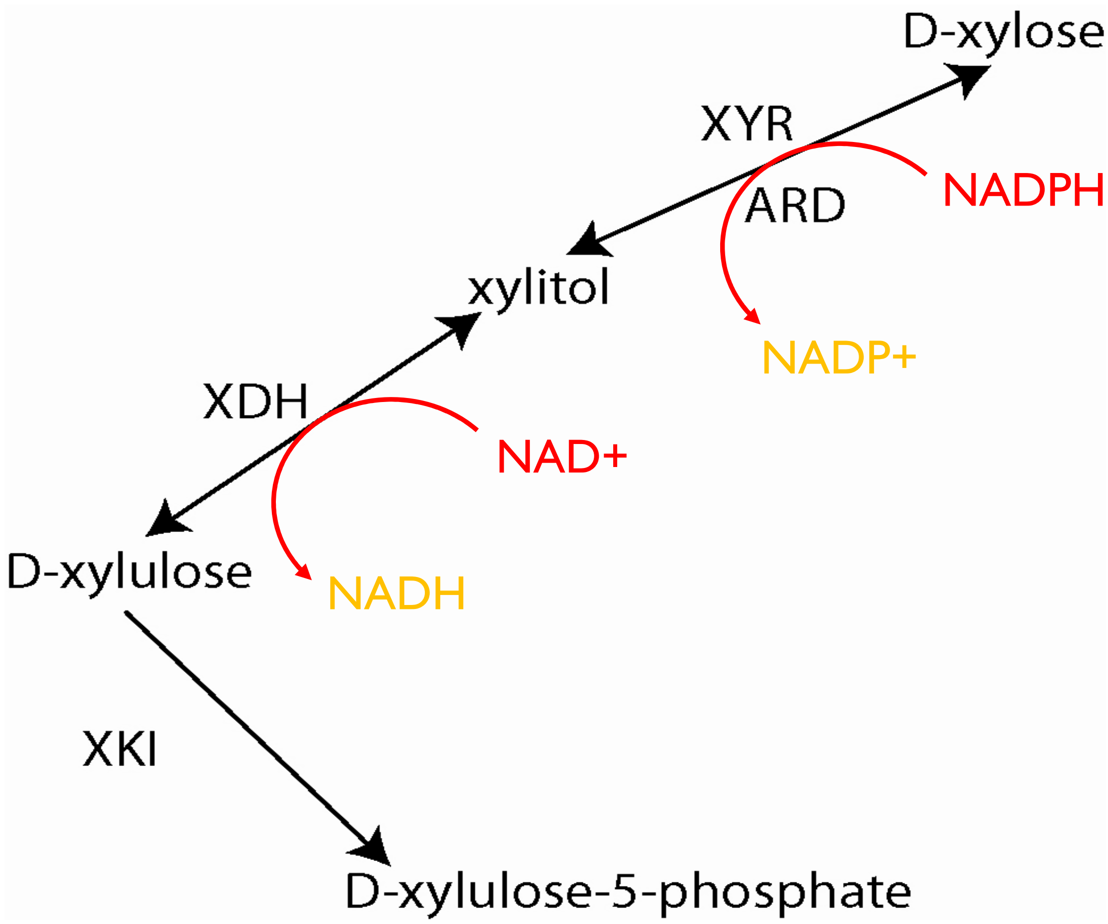
CAN CODA BE USED TO SOLVE THESE PROBLEMS?

PENTOSE SUGAR UTILIZATION PATHWAYS

- FUNGAL

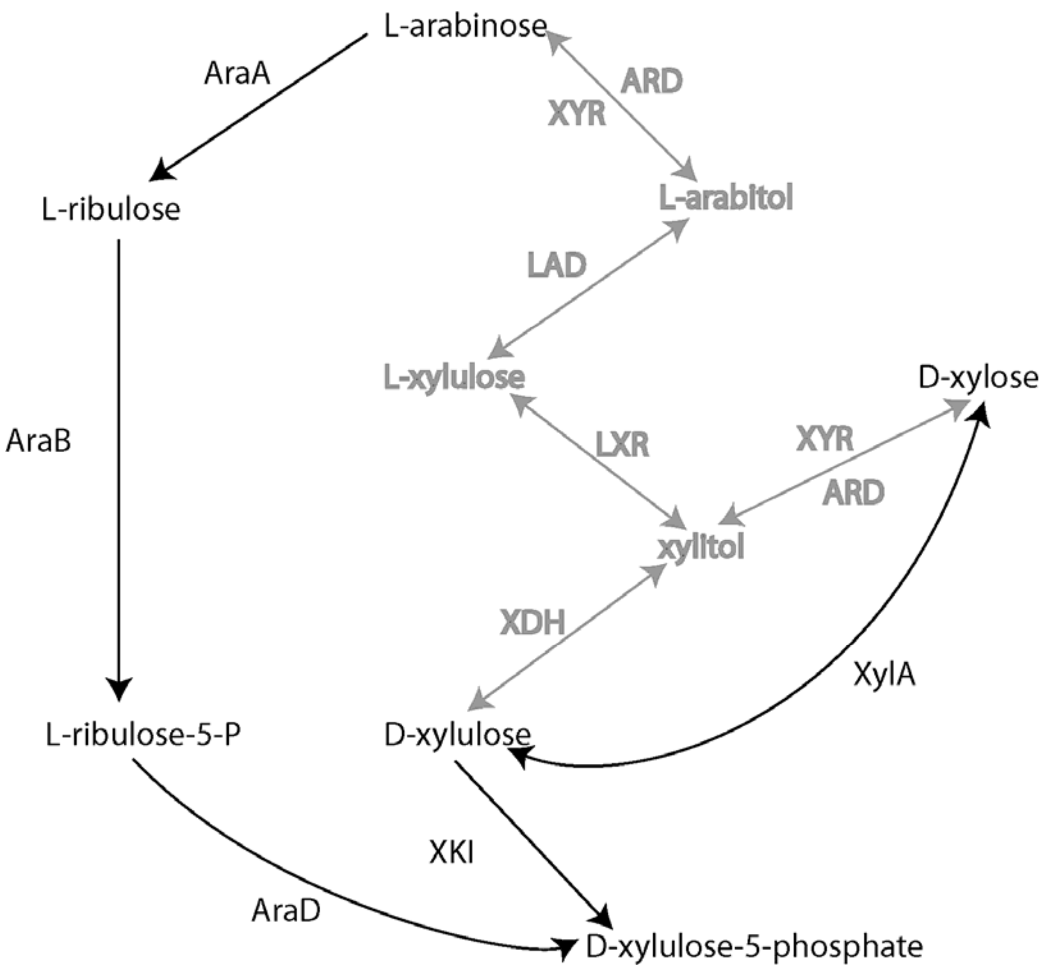


THE REDOX PROBLEM



PENTOSE SUGAR UTILIZATION PATHWAYS

FUNGAL
BACTERIAL

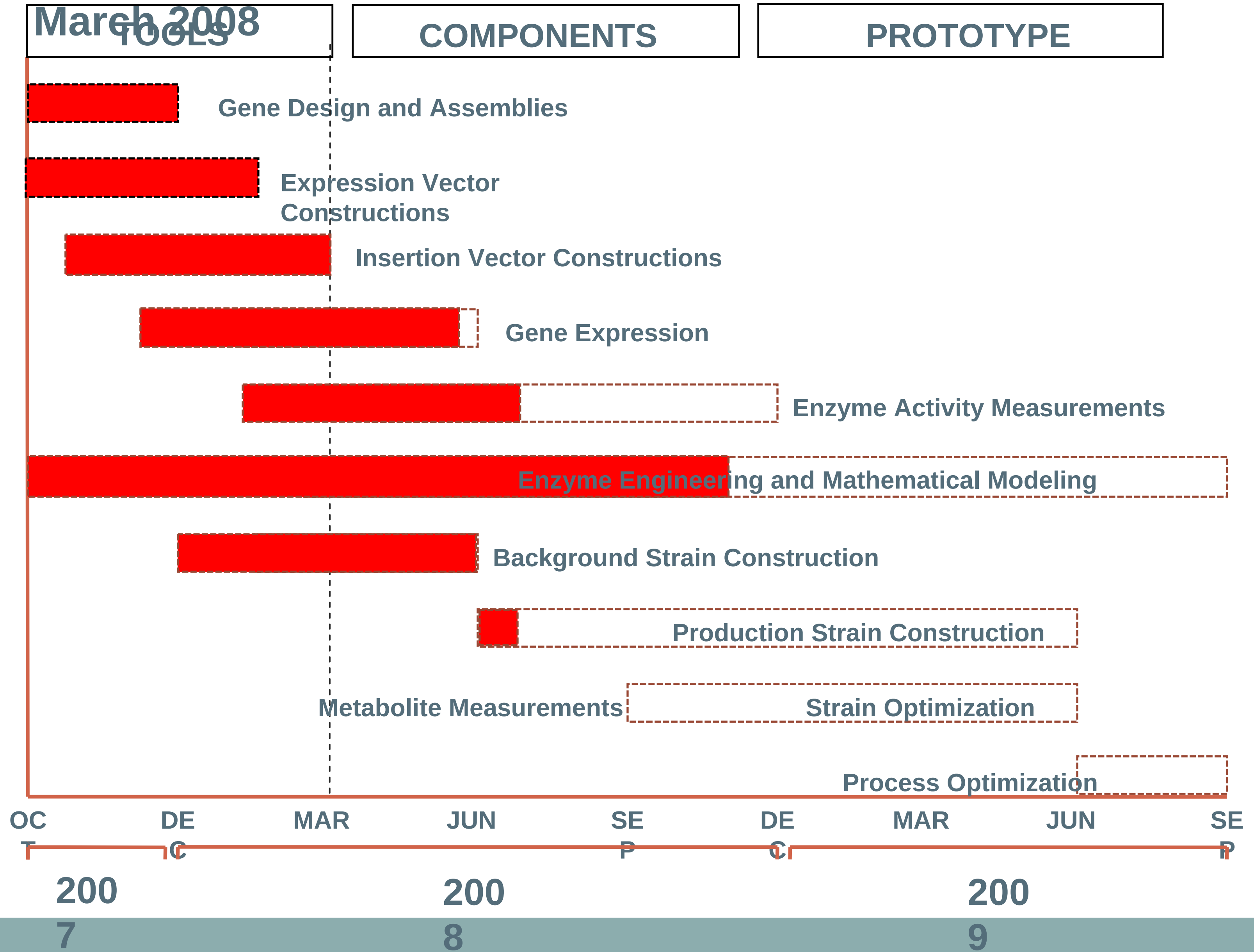


PROJECT STRATEGY

- CODA gene design, assembly and expression
  - Fungal pathways
  - Bacterial pathways
- Novel vector design and construction
- Enzyme purification and biochemical analysis
- Construction of a yeast production strain
- Strain optimization: math modeling; protein engineering; directed evolution; genomics

UCI/CODA PROJECT PROGRESS

March 2008



THE UCI BIOFUELS TEAM

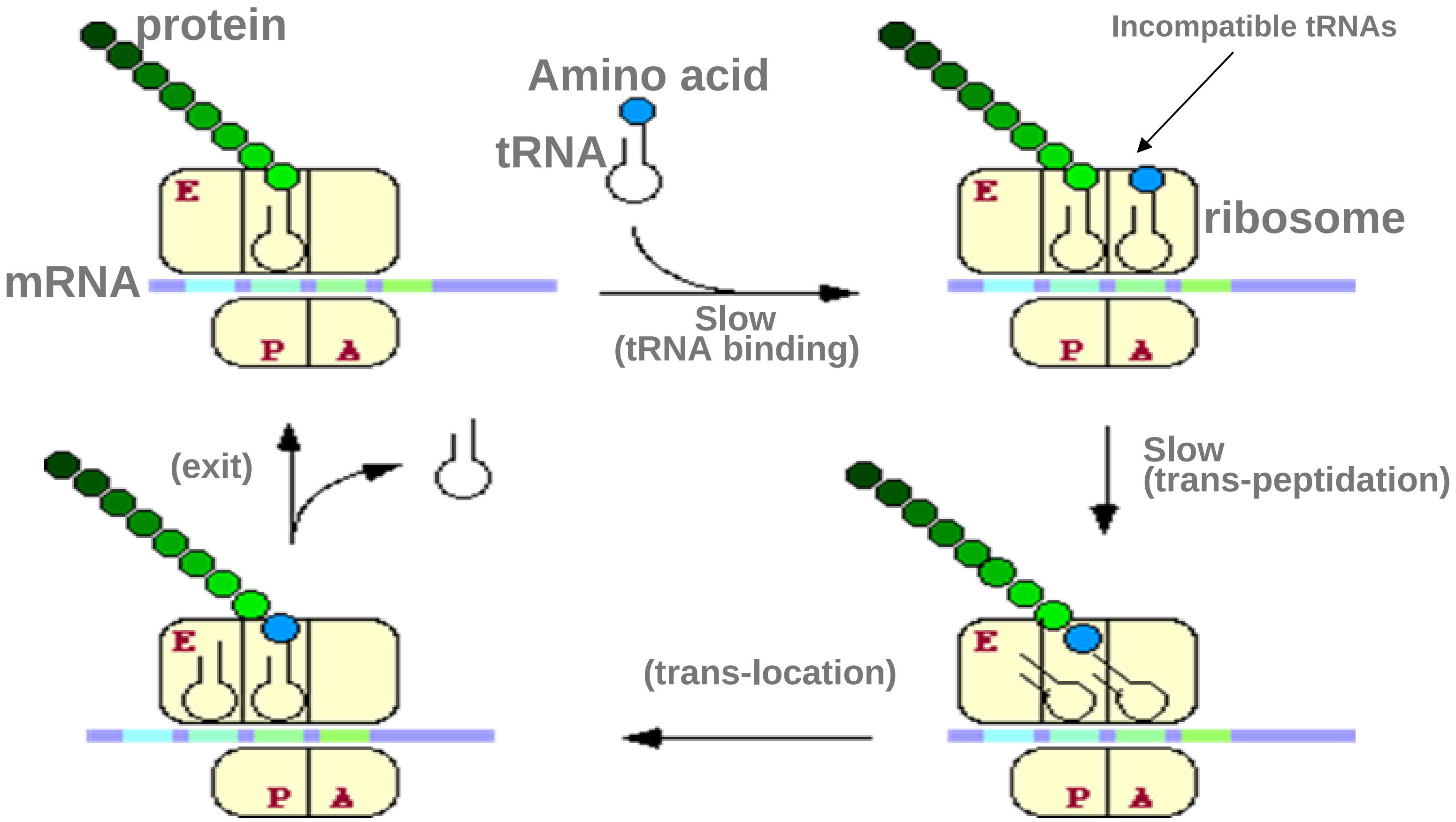
- **Mol. Biol. Biochem.**
  - **Wes Hatfield, Ph.D.**
  - Kirsty A. Salmon, Ph.D.
  - Kimberly Aeling, Ph.D.
  - Elaine Ito, M.S.
  - Lydia Ho, B.S.
  - Alex Lyon
- **Computation**
  - **Pierre Baldi, Ph.D.**
  - Tarek Najdi, Ph.D.
  - Arlo Randall, Ph.D.
- **Chem. Eng.**
  - **Nancy Da Silva, Ph.D.**
  - Jin Wook Choi, M.S.
  - Sneha Srikrishnan, B.S.
  - Mike Shen, M.S.
- **Yeast Mol. Biol.**
  - **Suzanne Sandmeyer, Ph.D.**
  - Uyen Tran, Ph.D.
  - Becky Irwin, B.S.

CODA OPTIMIZES GENE SEQUENCES FOR:

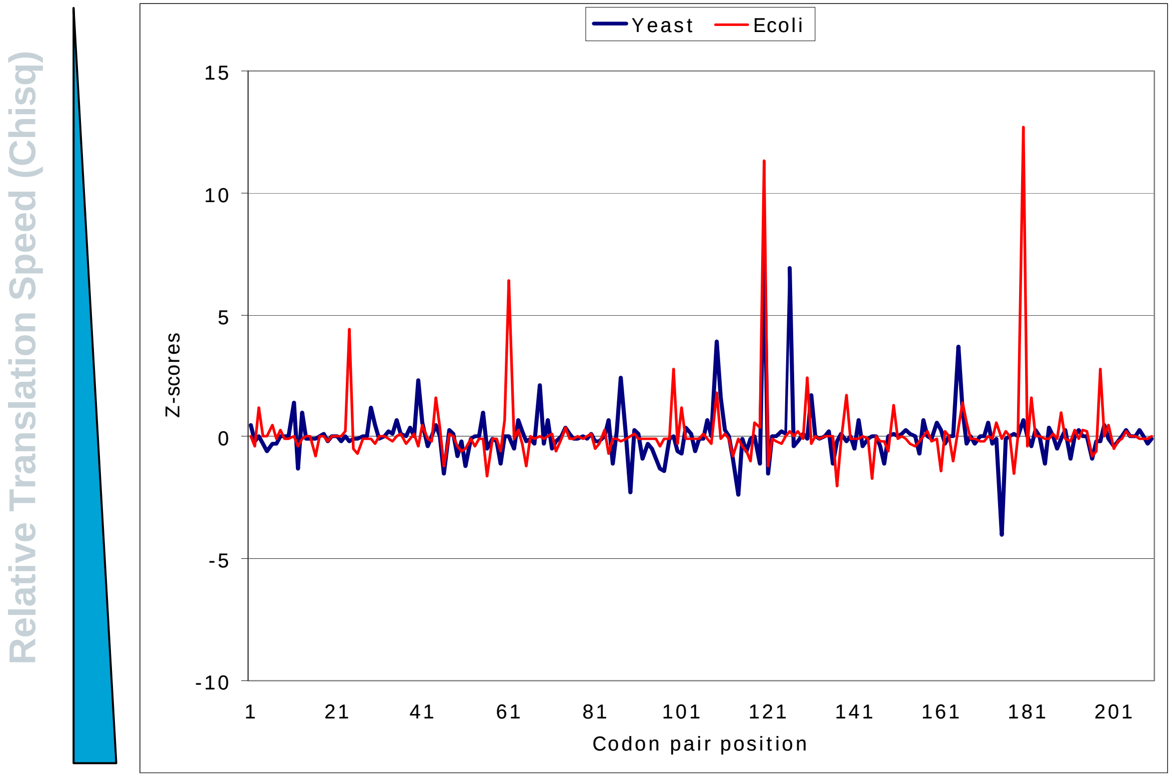
- * minimum melting temperature gap (CODA patent)
- * minimum average codon usage
- * maximum average codon pair chi-square (z score)
- * minimum absolute codon usage
- * maximum absolute codon pair (z score)
- * minimum maximum usage in adjacent codons
- * no Shine-Delgarno sequence (E. coli)
- * no occurrences of 5 consecutive G's or 5 consecutive C's
- * no long exactly repeated subsequences
- * no cloning restriction sites
- * no user-prohibited sequences (e.g., other restriction sites)
- * (optionally) no codon usage of a specific codon above user-specified limit
- * prefer high average codon usage
- * prefer low average codon pair chi-squared
- * prefer larger melting temperature gap
- * prefer more or fewer out of frame STOP codons (framecatchers)
- * (optionally) prefer evenly distributed codon usage

Codon Pairs and Translational Pauses

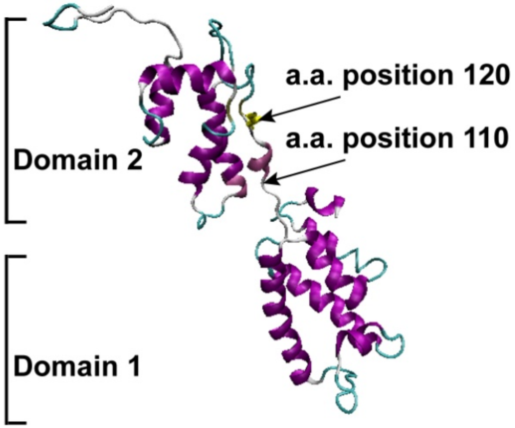
Over-represented codon pairs (incompatible tRNAs) are translated slower than under-represented pairs and define translation pause sites



SpeedPlot of Yeast Ty3 GAG gene in Yeast and *E. coli*



B.



HYPOTHESIS:

The frequency of one codon next to another has co-evolved with the structure and abundance of tRNA isoacceptors in order to control the rates of translational step times without imposing additional constraints on amino acid sequences or protein structures

DOWNSTREAM PATHWAYS