# INFORMATION PROCESSING

## **20.** Intracellular Errors.

Transcript fidelity

Translational fidelity.

Biophysics of substrate discrimination and the cost of proofreading.

The limits to selection on error rates.

The evolutionary consequences surveillance-mechanism layering.

Adaptive significance of errors.

#### 21. Intracellular Communication: Transcription.

Transcription factors and their regulatory motifs.

Biophysics of recognition.

Facilitated diffusion and the search for regulatory motifs.

Evolution of the regulatory vocabulary.

Evolutionary rewiring of transcription networks.

RNA-mediated expression.

#### 22. Environmental Sensing and Extracellular Communication.

Bacterial signal-transduction systems: origin; coevolution of components; emergence of new pathways. Interconvertible proteins and ultrasensitivity; and the cost of signal transduction.

Chemotaxis; and the accuracy of environmental assessment.

Phenotypic bimodality and bet-hedging: adaptive fine-tuning vs. inadvertent by-products of pathway structure.





The only trait for which we have a comprehensive theory for the evolution of mean phenotypes across the Tree of Life.  In eukaryotes, Pol II is reserved for the production of messenger RNAs and micro RNAs; Pol I for the synthesis of ribosomal RNA subunits; and Pol III for transfer RNA production. Land plants deploy two additional RNA polymerases to generate small RNAs used in transcriptional silencing (Wierzbicki et al. 2009; Haag and Pikaard 2011; Werner and Grohmann 2011).



- The complexity of these enzymes is quite variable:
  - Bacterial and archaeal RNAPs consist of 5 and 12 subunits respectively.
  - Eukaryotic Pols I and III contain 14 and 17 subunits respectively, and Pols II, IV, and V contain 12 subunits.
- Yet, despite the fact that eukaryotic RNA polymerases are more complex than those from bacteria, there is no evidence that the former carry out their tasks more efficiently or more rapidly.
- Transcription is slower than replication in prokaryotes, whereas both processes proceed at similar rates in eukaryotes.
  - Transcription rates (bp / sec) are: ~46 in *E. coli*, 20 to 60 in yeast, 21 in *Drosophila*, and 56 in human.
  - Replication rates are in the range of 100 to 1000 bp / sec in prokaryotes, but 10 to 50 bp / sec in yeast, flies, and mammals.



Sort into uniquely barcoded families; search for consistent errors:







Stephan Baehr

Weiyi Li





Jean-Francois Gout

Marc Vermulst

- Interrogation of ~2.5 billion bp yielded >200,000 transcription errors in 8 cell lines.
- Base-substitution error rate =  $4 \times 10^{-6}$  / site, independent of gene-expression level.







• Only a 5-fold range of variation across the Tree of Life.

- The transcript-error rate is function of both the population and cellular environments.
- Eukaryotic species have increased proteome sizes and average gene lengths, increasing the target size for errors.
- Larger eukaryotic cells harbor more transcripts per gene, and if the effects of errors on fitness are synergistic, the net load on cell fitness will be increased in larger cells, increasing the strength of selection for transcript fidelity.





Dilution effect ( << 1.0 );  $f(n) = (1/n)^x$ 



- P = number of protein-coding genes / proteome
- L = mean length of protein
- n = mean number of transcripts / gene
- $N_e$  = effective population size



Human data: Milholland et al. 2017; Bae et al. 2018; Lee-Six et al. 2018; Moore et al. 2019; Yokoyama et al 2019.

### **Translational Fidelity**



- Amino-acyl tRNA synthetase proteins (AARSs), each assigned to a single amino acid, must initially harvest cognate AAs.
- Charged AARSs must pass their cargo on to the appropriate transfer RNA (tRNA).
  - Considerable room for error in both steps because the structural differences between some AAs are quite minimal, e.g., valine and isoleucine differ by the presence of just a single methyl group.
  - Most AARSs have proof-reading mechanisms to minimize misloading errors, although some species of *Mycoplasma* have lost the capacity for proof-reading in multiple synthetases.
- At the ribosome, each codon in an mRNA must be recognized by its cognate tRNA via codon : anticodon recognition.
  - Proof-reading occurs twice after tRNA loading, involving processes that require GTP hydrolysis.

• Using target genes devoid of codons for a particular amino acid, and then searching for the incidence of that amino acid in synthesized proteins.

• Using genes engineered to produce defective products unless a particular codon is misloaded by a specific amino acid, with the degree of rescue providing insight into the specific error rate at that one codon.

Misreading of a termination codon as a sense codon – monitoring the expression of reporter constructs containing
premature termination codons that completely abrogate gene function unless experiencing read-through.

• Mass spectrometry of individual protein fragments.

• The translation-error rate per codon is ~10<sup>3</sup> to 10<sup>4</sup> times greater than the transcript-error rate per nucleotide site.



$$E + R \xrightarrow{k_{a,R}} ER \xrightarrow{k_{cat,R}} E + P_R$$

$$E + W \xrightarrow{k_{a,W}} EW \xrightarrow{k_{cat,W}} E + P_W$$

R, W = "right" and "wrong" substrate molecules

• Error rate  $\approx$  ratio of dissociation rates of R vs. W

$$\frac{k_{\rm d,R}}{k_{\rm d,W}} \simeq e^{-\Delta E/(k_B T)}$$

- Most enzymes bind their specific substrates with energies in the range of 12 to 24  $k_BT$ .
  - The strength of a single hydrogen bond is ~5 to 15 k<sub>B</sub>T, depending on the context.
  - G:C pair involves 3 hydrogen bonds; A:T involves 2.
- Letting 2 or 5  $k_B T$  be the binding-energy differences yields error rates of  $\approx 0.13$  and 0.007.
- Extreme differences of 10 and 15  $k_B^T$  still yield error rates of 5 x 10<sup>-5</sup> and 3 x 10<sup>-7</sup>.



 Exploits weak binding energies and even smaller differences between substrates to repeatedly interrogate bound substrates until passing them on to the next biochemical stage.

 In principle, no limits to the level of accuracy that can be achieved, but increased accuracy comes at the cost of increased reaction times and energy consumption.

- Hopfield et al. (1976) showed directly without any detailed knowledge of the underlying mechanism.
  - *In vitro* system involving the transfer of either isoleucine or valine to the isoleucine tRNA via isoleucyl-tRNA synthetase.
    - When isoleucine was the sole substrate, 1.6 ATP hydrolyses occurred per charged isoleucyl-tRNA (implies that a correct substrate molecule is examined ~1.6 times prior to permanent attachment).
    - When valine was the sole substrate, 270 ATPs were consumed per charged tRNA.
  - Assuming equal substrate concentrations, these results suggest an error rate of ~1.6/270 = 0.006 resulting from the differential rejection of the two residue types.

- 25 to 40 ATPs are consumed when properly charged AARSs are forced to deliver an amino acid to a noncognate tRNA, implicating energy-consuming proofreading at the tRNA stage (Yamane and Hopfield 1977).
- When properly charged tRNAs encounter inappropriate codons during translation, GTP is hydrolyzed, implicating additional proofreading at the codon-anticodon recognition step on the ribosome (Thompson and Stone 1977; Yates 1979).

• Intrinsic cost of interrogating correct substrate molecules ≈ 0.6 extra ATPs per incorporation.

 Extra cost of rejecting incorrect substrate molecules ≈ (1.6 / 270) errors / correct base x (270 - 1.6) additional ATPs per incorporation ≈ 1.6 additional ATPs.

• Total cost at amino-acid tRNA synthetase level  $\approx$  1 + 0.6 + 1.6 = 3.2 extra ATPs per incorporation,

~3x the cost without proofreading.

• Total fraction of a cell's energy budget  $\approx$  2 to 5%.



1) Nonsense-mediated decay

- 2) Non-stop decay
- 3) No-go decay
- Translation fidelity:
  - 1) Amino-acyl synthetase loading
  - 2) Transfer from AARS to tRNAs
  - 3) Codon / anticodon pairing



Effective selection for antimutators

**DRIFT BARRIER** 

Biased production of mutators

• Equilibrium mutation rate is expected to be negatively associated with the effective population size.



Mutation-rate classes

• How can a secondary layer of defense be added that breaks the drift barrier?

• If such a genomic addition is assimilated, what are the long-term consequences for the previous layer, the new layer, and the combined effects of both?

• Is biological "robustness" an indicator of adaptive progression, or is adaptation closer to a "zero-sum" game?



- Rapid improvement accompanies establishment of a new layer of protection.
- Both layers then gradually become less efficient.
- The level of overall performance returns to that for the single-layered state.

- The "Paradox of Robustness" (S. Frank, PLoS One): a more complex system evolves, but nothing is gained.
- Something has been lost: sensitivity of the system to mutational breakdown and energetic cost of constructing and operating has increased.



- Selection operates to drive the joint effects of two traits down to the limits imposed by drift.
- There is a ridge along which the population can freely drift, even to the extent of losing one trait.

| Total Error Rate = | polymerase error rate      |
|--------------------|----------------------------|
|                    | х                          |
|                    | proof-reading error rate   |
|                    | x                          |
|                    | mismatch-repair error rate |

- Bacterial species devoid of (*Mycoplasma smegmatis*) or with low capacity (*Deinococcus*) for MMR have mutation rates as low as found in species with high MMR efficiency.
- Experimental evolution experiments with *E. coli* with MMR deleted quickly evolve "wild-type" mutation rates.



- An extreme adaptationist argument mistranslation is a regulated phenomenon, with organisms ``deliberately'' making errors in order to expand the chemical diversity of the cell.
  - An intermediate level of mistranslation, fine-tuned by natural selection, yields populations of variant molecules, some of which will have large enough fitness-enhancing functions to offset the deleterious effects of others.

- Imposition of high translation-error rates leads to malfunctioning cells in *Salmonella*, a 10x increase in the error rate leads to a 2x reduction in cell division rates.
- Removal of mRNA surveillance mechanisms, such as NMD, causes substantial fitness loss.

- *E. coli* engineered to be auxotrophic (unable to synthesize a particular nutrient) by introducing a missense mutation in a gene required for the synthesis of the nutrient, can be partially rescued with an editing defective tRNA synthetase (Min et al. 2003).
  - Such an extreme starting point provides little (if any) evidence for the adaptive significance of error production, as auxotrophic mutants are expected to be rapidly purged from populations by natural selection except in cases where the nutrient is freely available in the environment.

- Cells under stress often have translation-error rates increased by 10 to 100x.
  - When cells are stressed, cellular functions go wrong, and there is no obvious reason why translation should be immune to pathological behavior.
  - Examples promoted as poster children of adaptive translation inaccuracy are highly idiosyncratic:
    - Excess incorporation of leucine and serine in *Candida*;
    - Excess incorporation of phenylalanine and leucine in *Mycoplasma*;
    - Excess incorporation of asparagine and aspartate in *Mycobacterium*;
    - Excess methionine in *E. coli*, yeast, and mammals.



Increasing effective population size

Accuracy