STEVE BELL
TANIA BAKER

MITx 7.28.3x Molecular Biology: RNA Processing and Translation

1. Translation I

1.1. Introduction

- central dogma: DNA —(replication)→ DNA; DNA —(transcription)→ RNA; RNA —(translation)→ PROTEIN
  - Watson/Crick base-pairing ↔ genetic code mediated by tRNAs
  - specificity relies on

- translation
  - mechanism conserved across all living organisms → translation has evolved exactly once
  - hypothesis: fundamental mechanism catalysed by RNA (NOT proteins)

RNA enzymes → "learned" how to put amino acids together → loss of ability / lack of elaboration of RNA enzymes + development of DNA

↑ expanded possibilities (4 nucleic acids vs. 20 amino acids)

1.2 Reading the mRNA in Translation

universal genetic code: 64 codons (stop codons do not encode AA)

- non-overlapping code: 5' AUGAAA 3' vs. 5' AUGAAA 3'
  - overlapping would be very restrictive in terms of which AAs can follow after one another
- special codons
  - start: AUG (also Met)
  - stop: UGA, UAG, UAA
  - can vary in some organisms, but regardless of sequence, always codes for Met
- possible 3 reading frames
  - 5' AUG|AAA|GCA|AUU|UUC|GUA|CUG|AAA|GGU|UGG|UGG|CGC|ACU|UCC|UGA 3'
    start K A I F V L K G W W R T S stop
  - 5' A|UGA|AAG|CAA|UUU|UCG|UAC|UGA|AAG|GUU|GGU|GGC|GCA|CUU|CCU|GA 3'
    stop K Q F S Y stop K V G G A L P
  - 5' AU|GAA|AGC|AAU|UUU|CGU|ACU|GAA|AGG|UUG|GUG|GCG|CAC|UUC|CUG|A 3'
    E S N F R T E R L V A H F L

ORF: Open Reading Frame

RNA sequence that has potential to be translated
- starts w/ start codon (not necessarily the first one)
- ends w/ end codon

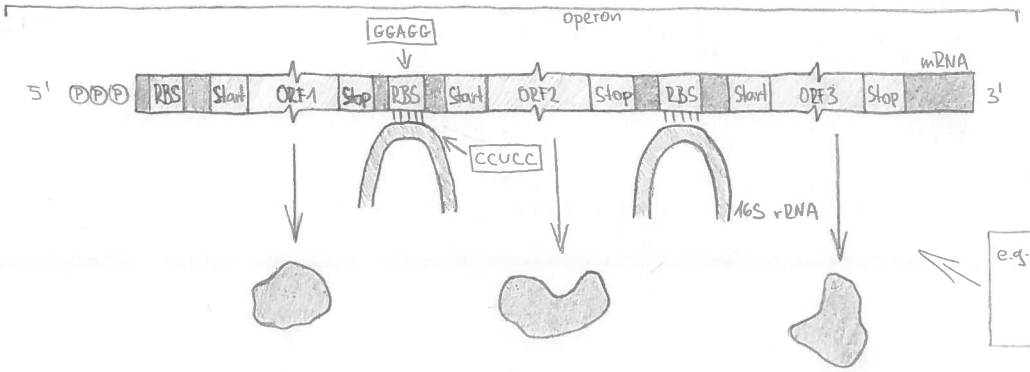
1.3 Bacterial mRNAs

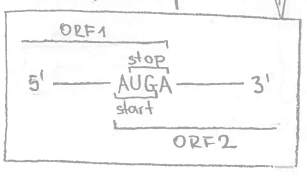
- can be polycistronic (can encode multiple proteins)
  - ORF ≃ cistron (historical, purely genetic construct: portion of RNA that encodes a protein)

mechanism for co-regulation of two (or more) proteins (eg. subunits of the same enzyme - need equal amounts of each)

polar effect: translation of ORF1 is required for translation of ORF2

- ORFs are NON-overlapping (except rare cases in viruses - need to keep genome as small as possible)
- RBS (ribosome binding site): GGAGG sequence 3-9nt upstream (5') of start codon — ribosome binding site = Shine-Delgarno site
  - recruits small ribosomal subunit by base-pairing w/ 16S rRNA (contains complementary sequence CCUCC)
  - ribosome recruitment independent of other RBSs on the mRNA (no need to bind/process upstream parts of mRNA)
  - no polar effect b/w individual RBSs (translation reinitiation immediately after termination, under certain circumstances)

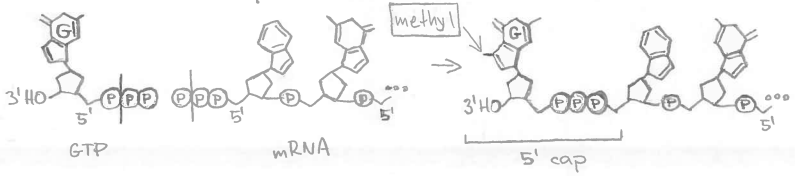




e.g. lac operon
- ORF1 = lacZ → β-galactosidase
- ORF2 = lacY → β-galactoside permease
- ORF3 = lacA → β-galactoside transacetylase

1.4 Eukaryotic mRNAs

- (almost) always monocistronic
  - exceptions: very small short proteins at the beginning of mRNA - no biological function (other than to regulate translation)
- no RBS present - there is consensus sequence (Kozak sequence): G/ANNNAUGGNN ← it promotes slightly higher translation levels, but is NOT essential (and often not present)
  (AUG = start)
- ribosome binds 5' cap of mRNA (see MITx 7.28.2x, pg. 9.1)



initiation machinery binds 5' cap

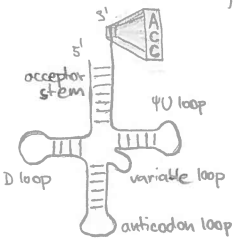
- ribosome scans mRNA for start codon (typically the first one, but there are exceptions - if AUG is too close to 5' end, it is hard to be recognized) (rare)
  - need to consider all 3 possible reading frames (as opposed to bacterial translation - RBS positions ribosome correctly) ←

?: 3-nt b/w RBS and AUG does not seem to be 1nt-precision, which is required

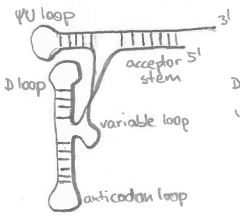
1.5 Structural Properties of tRNAs

- conserved among all organisms - it is very hard to make a mutation, because it influences many proteins at once
- structure
  - acceptor arm
  - ΨU loop
  - variable loop
  - anticodon loop
  - D loop
  - conserved: allow tRNA to be used in process of translation
    - ACC at 3' end
    - conserved bases
      - ΨU loop: 7
      - D loop: 4
      - anticodon loop: 1
    - # of bps in arms - D, ΨU, acceptor
  - distinct: different amino acids need to be charged to different tRNAs
    - anticodon
    - variable loop size
      - Class I: 3-5 bases
      - Class II: 13-21 bases
    - discriminator base: 4th base at 3'end (ACCD)
      - changes in discriminator base can have dramatic effects on ability to charge an amino acid to RNA

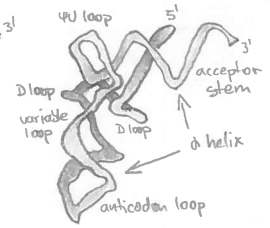
done by aminoacyl-tRNA synthetases (aa-tRNA synthetases) - critical role in translation fidelity



"clover leaf" schematic



real shape schematic



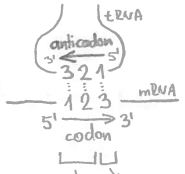
real shape crystal structure

- anticodon and amino acid are at opposite sides of tRNA molecule, far apart of each other
- each of them needs to be located in different part of ribosome during translation
  anticodon: small subunit
  amino acid: large subunit

1.6. The Genetic Code

- reconstructed by Har Gobind Khorana and Marshall W. Nirenberg
  - mRNA-free cell extract + raised Mg concentration to ~15mmol (no specific information for initiation is no longer needed) + artificial mRNA
  - observe protein produced
    - 5'-UUU...UUU-3' → Phe-Phe-...-Phe
    - 5'-CCC...CCC-3' → Pro-Pro-...-Pro
    - 5'-AAA...AAA-3' → Lys-Lys-...-Lys
    - 5'-GCG...GGG-3' → did not work, because poly-G RNA makes triple-stranded structure that can not be translated (physically impossible)
    - 5'-UCU...CUC-3' → Ser-Leu-...-Ser-Leu (UCU/CUC → Ser/Leu, but not enough information to tell which one is which)
    - 5'-ACA...CAC-3' → Thr-His-...-Thr-His (~11~)
    - more complex RNAs, many PhD students, ~10 years
    - stop codons determined by different sized protein products

1.7. Wobble Base Pairing



anticodon 1		codon 3
C	-	G
A	-	U
U	-	A/G (purine)
G	-	C/U (pyrimidine)
I (inosine)	-	A/C/U

e.g. both GAC and GAU code for Asp - single tRNA w/ anticodon GUC is sufficient
all three AUA, AUC and AUU code for Ile - single tRNA w/ anticodon IAU is sufficient

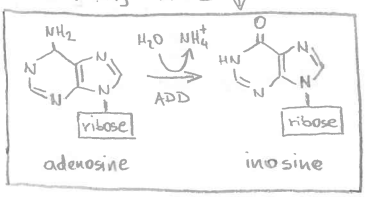
! It is NOT possible to recognize A and C independently
only as A/G only as C/U

UGC/U → Cys
UGG → Trp
UGA → stop
it is not possible to have tRNA specific for UGA; there is NO tRNA for stop codons

3: wobble base pair: different base pairing (relaxed, in comparison w/ W-C base pairing)
1,2: normal Watson-Crick bps

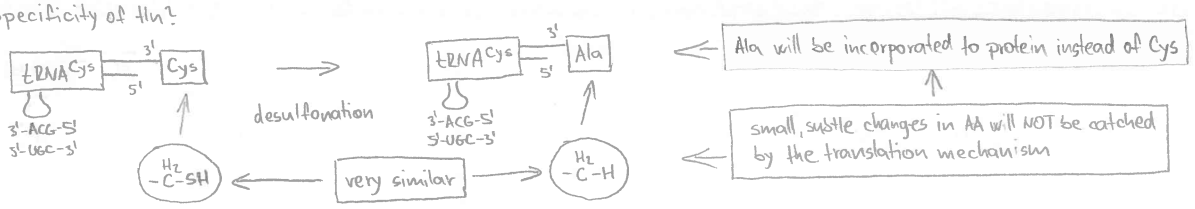
1.7.5 Inosine

- adenosine deaminases: deaminate A in already transcribed tRNA, transforming A into I
  - ADATs (Adenosine Deaminases for tRNA) - act on tRNA
  - ADARs (Adenosine Deaminases for RNA) - act on mRNA

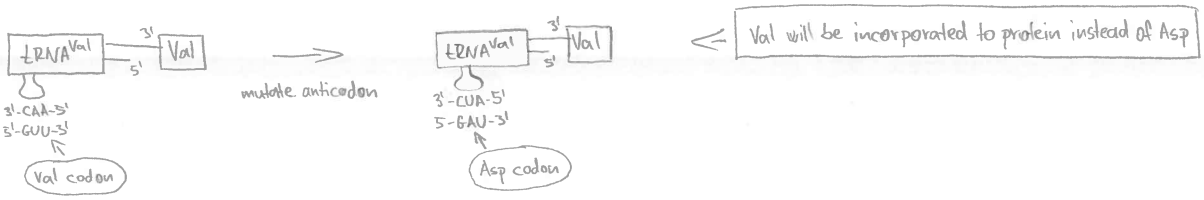


1.8. Specificity of Translation

- what generates specificity of tln?
- experiment:



- experiment



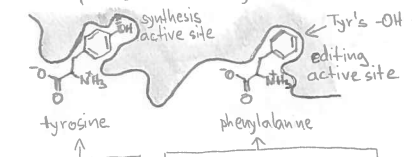
- correct charging of AAs on tRNAs is the most important for tln fidelity - done by ↓ ch. 1.9.

1.9. Aminoacyl-tRNA Synthetases

- mechanism:
  - AA + ATP → AMP-AA + PP_i ← amino acid adenylation
    - PP_i: rapidly degraded by pyrophosphatase
    - this bond (b/w AA and AMP) retains high energy nature (critical to drive tln) !!!
      - as opposed to DNA/RNA, there is NO triphosphate bond on AAs, but something still need to drive the reaction (tln)
  - AMP-AA + 5'-tRNA-CCA-3' → 5'-tRNA-CCA-AA + AMP ← tRNA charging
    - charged tRNA
- structure: relatively complicated (4 domains, 3 active sites)
  - acceptor stem binding domain: part of AA coupling active site
  - anticodon binding domain: only single base recognized (which of the three it is varies b/w individual AA tRNA synthetases)
  - ATP + AA binding domain: part of adenylation active site
  - AA hydrolysis domain: proofreads either (AA adenylation / tRNA charging) } varies b/w individual AA tRNA synthetases
    - proofreading active site
- single AA tRNA syntetase per AA
- recognized parts of tRNA
  - anticodon (typically only single base)
  - acceptor arm (1-3 bases + discriminator base)
  - other unique modifications / specific sequences

1.10 Aminoacyl-tRNA Synthetase Proofreading

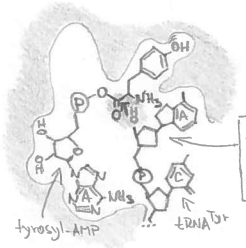
- proof reading is typically "designed" to find closely related AAs (most likely scenario of mistake: mischarging similar AA)
- example: tyrosil tRNA synthetase



- binds correct AA
- binds incorrect AA does NOT bind correct AA (steric clash w/ tyrosine)

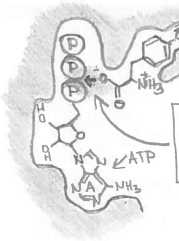
← chemical proofreading

- proofreading mechanisms
  - chemical proofreading: editing active sites chemically cleaving incorrect AAs
  - kinetic proofreading: correct (AA, tRNA) pair allows correct positioning for catalysis (correct reaction is much faster)
    - similar to e.g. DNA synthesis, where positioning of correct NTP allows for much faster incorporation than positioning of incorrect NTP



- -OH of $tRNA^{Tyr}$ attacking carboxyl group of incomming adenylated AA (highlighted area)

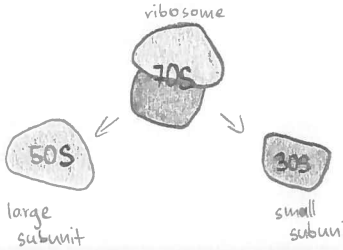
← kinetic proofreading



- carboxyl group of AA attacking alpha phosphate of ATP (highlighted area)

1.11 Bacterial Ribosome Structure Overview

- massive structure: ~2.5MDa (~5x bigger than RNA Polymerase holoenzyme - the biggest structure presented so far in MITx 7.28)



S (Svedbergs)
- unit of relative sedimentation velocity
- named after inventor of centrifuge
- historically used to measure relative sizes of ribosome and its subunit
- although the unit is not used anymore, historical names of subunits stick

- composition: ribonucleoprotein (50% RNA, 50% protein)
  - large subunit
    - 5S RNA (120 nt)
    - 23S RNA (2900 nt)
    - ~34 proteins
  - small subunit
    - 16S RNA (1540 nt)
    - 21 proteins

<=
- proteins
  - not involved in enzymology of the reaction (catalysis)
  - only hold and help correct folding of rRNAs
- RNAs - mediate catalysis (ribosome = rybosime => RNA catalysis)

1.12. Bacterial Crystal Structure of the Ribosome

- core mostly made of RNA
- proteins tend to be on the outside of the structure
- channels
  - mRNA/tRNA channel
  - peptide exit channel (or tunnel): too tight for peptide folding to occur inside the channel

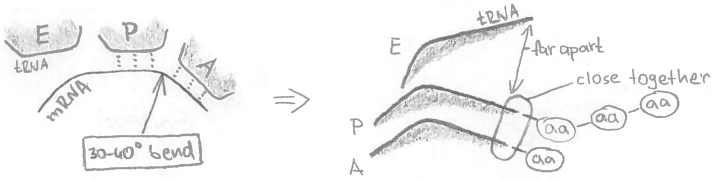
only α-helices are formed inside the channel
even β-sheets are too large to fit in

1.13. Key Functional Regions in the Ribosome Structure

- three tRNA binding sites
  - A - binds aminoacyl tRNAs
  - P - binds peptidyl tRNA (nascent peptide bound)
  - E - binds uncharged tRNA (Exit, Expended, ...)
- structural features
  - A,P tRNAs interacting w/ mRNA
  - E tRNA NOT interacting w/ mRNA
  - mRNA kinked b/w A and P sites in 30-40° angle
    - prevents tRNA from accidentaly binding nts it is not supposed to bind
    - enforces correct reading frame (prevents frame shifts)

decoding center

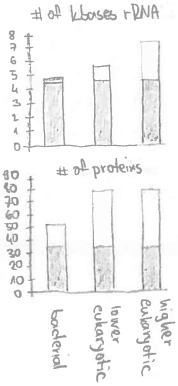
peptidil transferase center
- acceptor ends of A and P close to one another
- acceptor end of E far apart
- reaction occurs b/w A and P

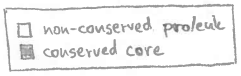


decoding center (small subunit) | peptidil transferase center (large subunit)

1.14. Bacterial versus Eukaryotic Ribosomes

	bacterial	eukaryotic
whole ribosome	70S (2.5MDa)	80S (4.2MDa)
large subunit	50S (1.6MDa)	60S (2.8MDa)
rRNAs	5S (120 nt) 23S (2900 nt)	5S (120 nt) 5.8S (160 nt) 28S (4700 nt)
proteins	~34	49
small subunit	30S (0.9MDa)	40S (1.4MDa)
rRNAs	16S (1540 nt)	18S (1900 nt)
proteins	21	~33





- core region - almost absolutely conserved in ALL organisms
- proteins mostly on the outside of ribosome
- enzymatic core made of rRNA

proteins provide
- structure
- regulation