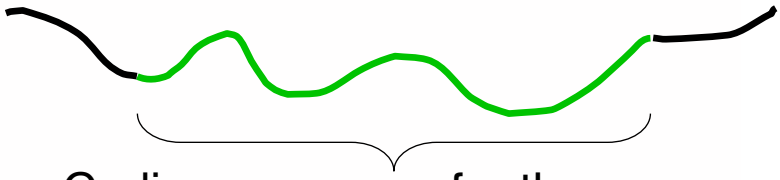
Biomolecular chemistry

3. Translating the genetic code

Primary Source Material

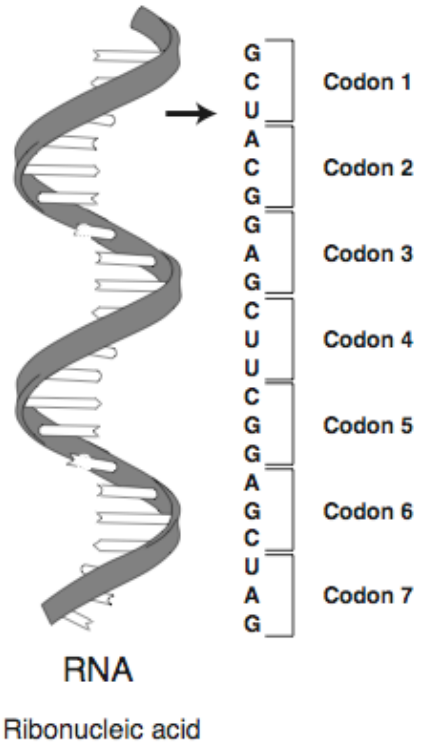
- Chapter 29 of Biochemistry Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert (courtesy of the NCBI bookshelf)
- Molecular Cell Biology Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. (courtesy of the NCBI bookshelf)
- Many figures and the descriptions for the figures are from the educational resources provided at the Protein Data Bank (http://www.pdb.org/)
- Most of these figures and accompanying legends have been written by David S. Goodsell of the Scripps Research Institute and are being used with permission. I highly recommend browsing the Molecule of the Month series at the PDB (http://www.pdb.org/pdb/101/motm_archive.do)

The genetic code



Coding sequence for the green fluorescent protein (GFP)

CACGCCUGCAGGUCGACUCUAGAGGAUCCCCGGGUACCGGUCGCCACC**AUGGUGAGCAAG**
GGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAAC
GGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACC
CUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCACC
CUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUC
UUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGACGAC
GGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUC
GAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGGCACAAGCUGGAGUAC
AACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUG
AACUUCAAGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAG
CAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACC
CAGUCCGCCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUC
GUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAAAGCGGCCGCGAC



◦AccessExcellence: The National Health Museum Graphics Gallery

- The genetic code is the relation between the sequence of bases in DNA (or its RNA transcripts) and the sequence of amino acids in proteins. Experiments by Francis Crick, Sydney Brenner, and others established the following features of the genetic code by 1961:
- Three nucleotides encode an amino acid. Proteins are built from a basic set of 20 amino acids. Simple calculations show that a minimum of three bases is required to encode 20 amino acids. An amino acid is encoded by a group of three bases, or codon.
- The code is nonoverlapping. Consider a base sequence ABCDEF. In an overlapping code, ABC specifies the first amino acid, BCD the next, CDE the next, and so on. In a nonoverlapping code, ABC designates the first amino acid, DEF the second, and so forth.
- The code has no punctuation. The sequence of bases is read sequentially from a fixed starting point, with no interruptions.
- The genetic code is degenerate. Some amino acids are encoded by more than one codon (there are 64 possible base triplets and only 20 amino acids). In fact, 61 of the 64 possible triplets specify particular amino acids and 3 triplets (called stop codons) designate the termination of translation. Thus, for most amino acids, there is more than one code word.
- Note that the strand of DNA which is read by RNA polymerase is known as the template strand. The complement of the template strand is known as the coding strand (also known as the sense strand, because it makes sense in terms of the protein sequence). The coding strand (sense strand) has the same sequence as the resulting RNA (with Ts switched for Us of course).
- Why did nature pick 3 bp/codon? Why not 1,2,4 or…?

The genetic code

First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Note: This table identifies the amino acid encoded by each triplet. For example, the codon 5′ AUG 3′ on mRNA specifies methionine, whereas CAU specifies histidine. UAA, UAG, and UGA are termination signals. AUG is part of the initiation signal, in addition to coding for internal methionine residues.

- Because the code is highly degenerate, only tryptophan (Trp) and methionine (Met) are encoded by just one codon each. The other 18 amino acids are each encoded by two or more. Indeed, leucine (Leu), arginine (Arg), and serine (Ser) are specified by six codons each.
- Codons that specify the same amino acid are called synonyms. For example, CAU and CAC are synonyms for histidine. Note that synonyms are not distributed haphazardly throughout the genetic code. Most synonyms differ only in the last base of the codon.
- What is the biological significance of the extensive degeneracy of the genetic code? If the code were not degenerate, 20 codons would designate amino acids and 44 would lead to chain termination. The probability of mutating to chain termination would therefore be much higher with a nondegenerate code. Chain-termination mutations usually lead to inactive proteins, whereas substitutions of one amino acid for another are usually rather harmless (though could often be harmful and sometimes beneficial). Thus, degeneracy minimizes the deleterious effects of mutations.
- Degeneracy of the code may also be significant in permitting DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA. The G + C content of bacterial DNA ranges from less than 30% to more than 70%. DNA molecules with quite different G + C contents could encode the same proteins if different synonyms of the genetic code were consistently used.
- *Q: The other question is that if we have to memorize the 3-letter DNA/RNA codes for amino acids and their chemical structure or not?*
- *A: I will give you this table of codons on a test if you need it. You do need to know the 1 and 3 letter codes for the amino acids and their structures.*

Translated sequence for GFP

```
cacgcctgcaggtcgactctagaggatccccgggtaccggtcgccaccatggtgagcaag
                                                 M  V  S  K
ggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaac
 G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N
ggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacc
 G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T
ctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccacc
 L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T
ctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttc
 L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F
ttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgac
 F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D
ggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatc
 G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I
gagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtac
 E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y
aactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtg
 N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V
aacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccag
 N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q
cagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacc
 Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T
cagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttc
 Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F
gtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaagcggccgcgac
 V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
```

- In this translated sequence the top row (cacgcc..) and every second row following, is the original DNA sequence. The RNA sequence would have ‘u’ instead of ‘t’.
- The second row (MVSK…) and every second row following is the translated protein sequence. The one letter code for each of the 20 amino acids is used. M is used for methionine.
- Note that protein synthesis is initiated at the first AUG (ATG in this translation from DNA).
- Protein synthesis stops when the ribosome encounters the codon uaa (taa in this translation from DNA).
- The easiest way to do a translation is to use online software programs. My favourite is the Translate program at ExPASy (http://web.expasy.org/translate/)
- *Q: what is open reading frame (ORF)?*
- *A: This is another way of referring to the region of a gene that actually becomes a protein. Basically, it is the part of the gene that is not a 5' or 3' UTR and not an intron.*

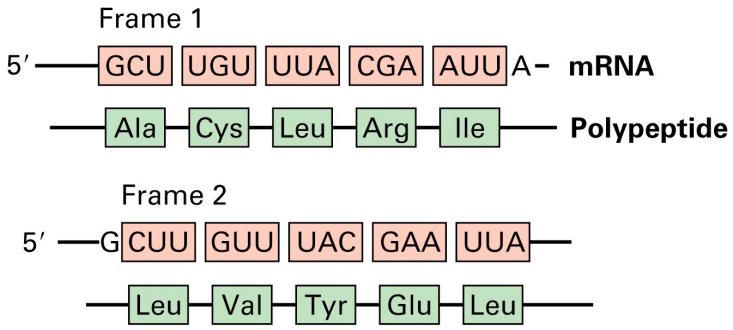
Having the correct reading frame is important: a metaphor

Imagine that all words in the English language were 3 letters long and there were no spaces or punctuation allowed. A typical sentence might read like:

REXTHEDOGHADONELEGBUTBOBTHEFOXHADTWO

```
You can only read this sentence if you know where to start
       +1 : REX THE DOG HAD ONE LEG BUT BOB THE FOX HAD TWO
       +2 : R EXT HED OGH ADO NEL EGB UTB OBT HEF OXH ADT WO
       +3 : RE XTH EDO GHA DON ELE GBU TBO BTH EFO XHA DTW O
```

The same is true of codons:



- So, the key to getting the right reading frame for a particular protein is correctly identifying the first set of 3 nucleotides that corresponds to the first amino acid of the protein. This first amino acid is always a methionine (Met), which is encoded by the AUG codon.

Having the correct reading frame is important: GFP example

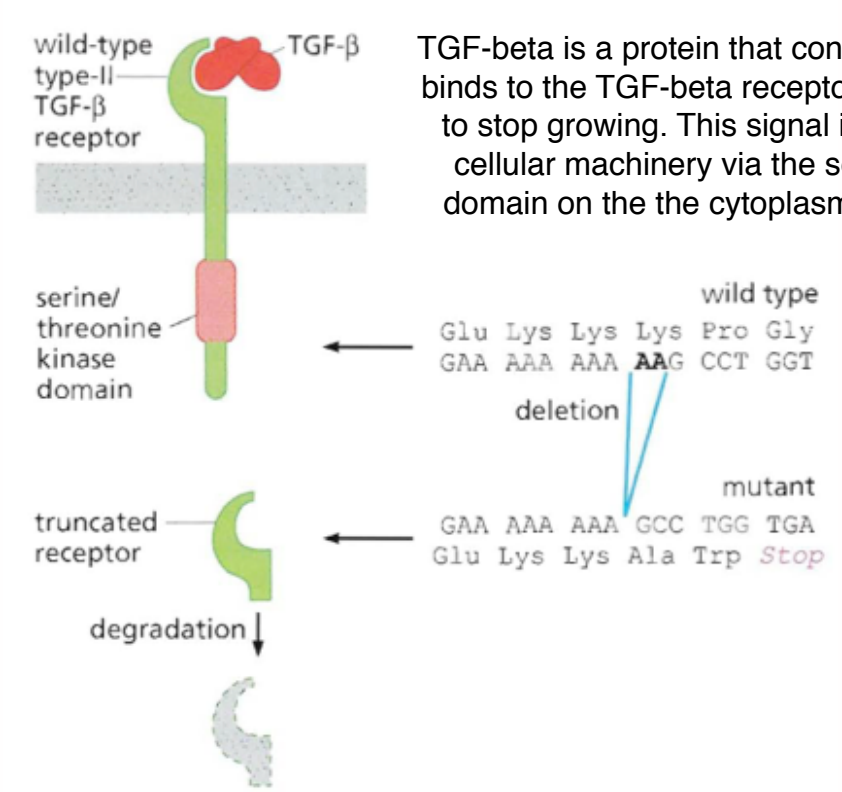
Consider the GFP gene:

```
RNA:  CACGCCUGCAGGUCGACUCUAGAGGAUCCCCGGGUACCGGUCGCCACCAUG
 +3:    R  L  Q  V  D  S  R  G  S  P  G  T  G  R  H  H  G
 +2:   T  P  A  G  R  L  *  R  I  P  G  Y  R  S  P  P  W
 +1: H  A  C  R  S  T  L  E  D  P  R  V  P  V  A  T  M
RNA:  GUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAG
 +3:    E  Q  G  R  G  A  V  H  R  G  G  A  H  P  G  R  A
 +2:   *  A  R  A  R  S  C  S  P  G  W  C  P  S  W  S  S
 +1: V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E
RNA:  CUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAG
 +3:    G  R  R  R  K  R  P  Q  V  Q  R  V  R  R  G  R  G
 +2:   W  T  A  T  *  T  A  T  S  S  A  C  P  A  R  A  R
 +1: L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E
RNA:  GGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGC
 +3:    R  C  H  L  R  Q  A  D  P  E  V  H  L  H  H  R  Q
 +2:   A  M  P  P  T  A  S  *  P  *  S  S  S  A  P  P  A
 +1: G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G
```

And so on….

- A protein translated in the wrong frame will be nonsensical
- If for some reason the start codon is missing, protein synthesis will start at the next AUG codon (we will often call it an ATG codon, which simply means that we are thinking about the gene at the DNA, as opposed to RNA, level) which may or may not be in the correct frame.
- If the next ATG encodes an in frame methionine, the translated protein will be missing the N-terminal sequence between its first two methionine residues.
- If the next ATG is out of frame (see +2 translation) then a nonsensical protein will be translated and a stop codon will likely be encountered soon after initiation.
- An insertion mutation within the gene for a protein can also cause the translation of nonsensical protein sequence.
- Insertions before or after the gene will not have an effect because only the sequence between the start codon (ATG) and the stop codon is relevant.
- *Q: if we are given a gene sequence in the order 5'-3' for example let it be 5' - AGC AAG GGC TAG CAG - 3' and we are asked to do the translation in the given soft ware we will be getting answers of different frames(5'-3') and also from (3'-5'). My question is what does this 3'-5' actually means? Does it mean that we look the given sequence in the opposite direction and if so why during translation from 3'-5' the basic bases are changing for example in the above case we get 3'-CTG CTA GCC CTT GCT -5'.*
- *A: I think that you are asking what the 3'-5' translation output provided by the software actually means. The most important thing to keep in mind is that we never write (or even need to think about) DNA in the 3' to 5' direction. So when the software provides the 3' to 5' translation it is actually providing the 5' to 3' translation of the reverse complement. It is a bit confusing that the software refers to it as a 3'-5' translation.*

Cancer can take advantage of “frame-shift” mutations

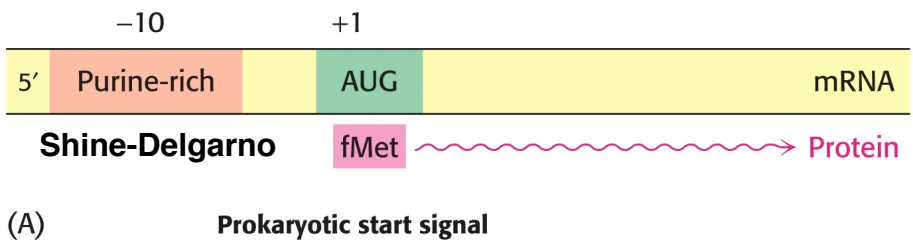


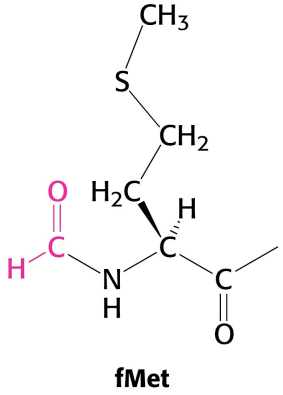
TGF-beta is a protein that controls cell growth. When it binds to the TGF-beta receptor, it is a signal to the cell to stop growing. This signal is communicated to the cellular machinery via the serine/threonine kinase domain on the the cytoplasmic side of the receptor.

Chapter 12 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

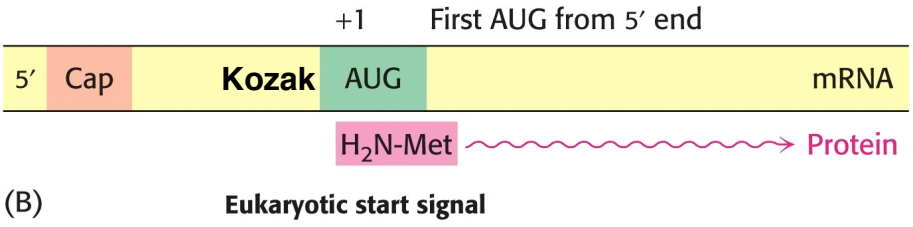
- The gene encoding the type II TGF-beta receptor is frequently inactivated in certain types of human colorectal cancers.
- TGF stands for “transforming growth factor”. It is a signal that a cell should not grow and divide. That is, it inhibits cell growth.
- It has been found that 90% of colorectal cancers have a defective TGF-beta receptor. In most cases, the defect is caused by the deletion of 2 (or more) “A”s from a region that normally has 10 As in row.
- Deleting 2 As causes a shift in the reading frame such that the subsequent sequence changes as a stop codon now occurs just a few codons later. This means that the protein is now 129 amino acids long, rather than the normal 565.
- This truncated protein is no longer functional and thus the cell has lost the ability to be inhibited by TGF-beta. Presumably, this gives the cancer cell a large growth advantage relative to other cells in the same tissue.

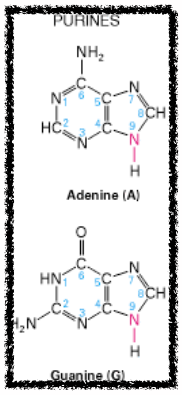
How does the ribosome know which AUG codon is the start codon?





fMet

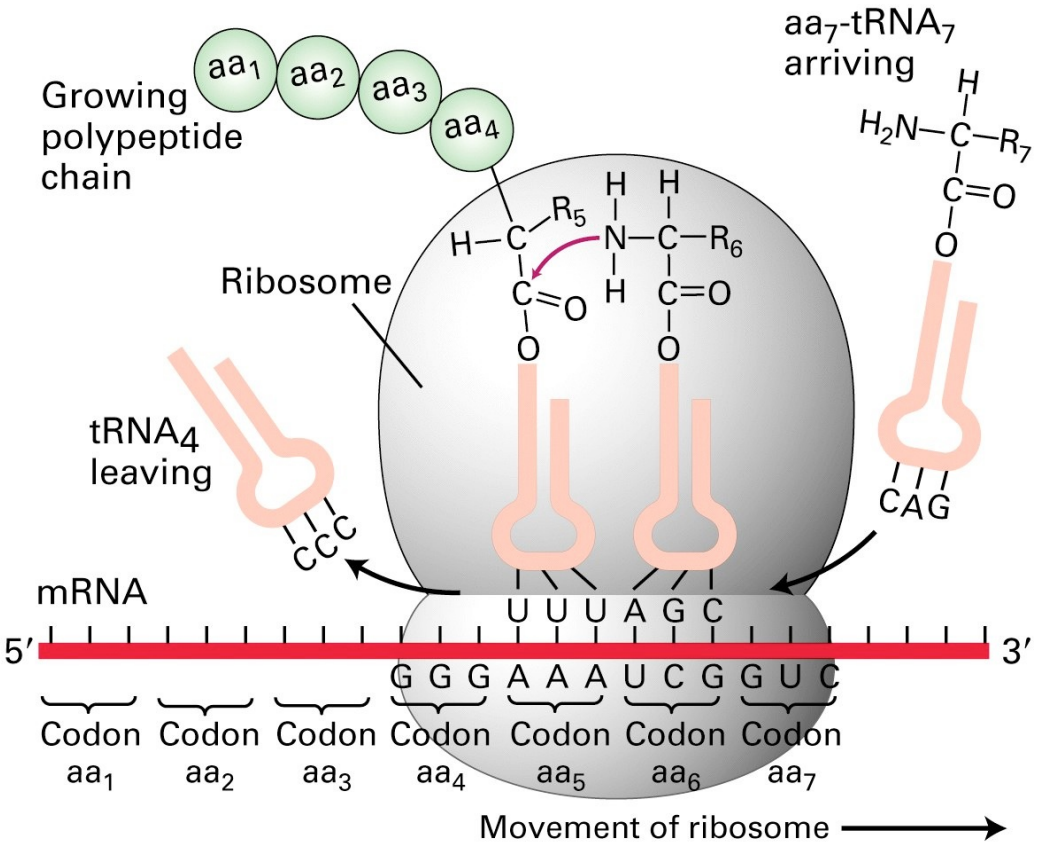




Is the gene from the previous example probably intended for bacterial or eukaryotic cell expression?

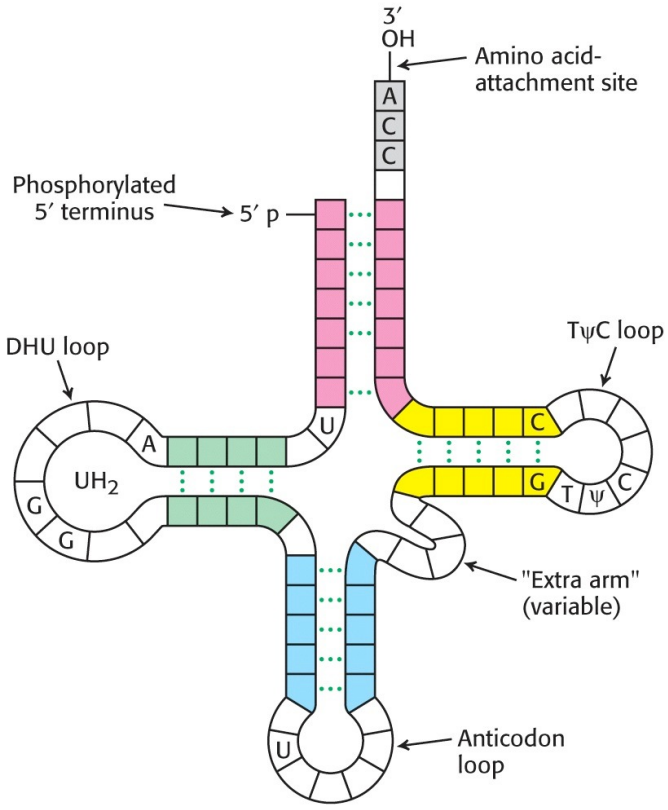
- How does the ribosome know where to start 'reading' the mRNA and translating the sequence into a polypeptide chain?
- In bacteria there is a specific tRNA, known as the initiator tRNA, that carries fMet. This fMet-tRNA recognizes the first codon AUG following a purine-rich sequence, known as the Shine-Delgarno sequence (or box), that base-pairs with a complementary sequence in the ribosome. This is essentially the attachment point for the ribosome onto the mRNA. The consensus sequence is AGGAGG (note that A and G are the purines).
- In eukaryotes, the 5' cap itself can serve as the ribosome binding site. The AUG closest to the 5' end of an mRNA molecule is usually the start signal for protein synthesis and this particular AUG is read by an initiator tRNA conjugated to methionine. The overall efficiency of translation also depends on the sequence immediately before the start codon. The optimal sequence is known as the Kozak consensus sequence. The optimal sequence is (gcc)gccRccAUGG, where R = A or G and the AUG portion is the start codon.
- An additional type of eukaryotic ribosome binding site is known as an 'internal ribosome entry site' or IRES. These are less common but are useful in some molecular biology applications.
- Once the initiator AUG is located, the reading frame is established—groups of three nonoverlapping nucleotides are defined, beginning with the initiator AUG codon.
- *Q: How do we know how many nucleotides are in the 3' UTR? In the homework, theres no sequence which tells us where transcription stops, so does it just continue until the end of the given DNA sequence?*
- *A: One mechanism (intrinsic termination) involves the formation of an mRNA stem loop structure that is more stable than the DNA/RNA hybrid. This causes the RNA to dissociate from the DNA and transcription stops.*

Overview of protein synthesis (translation)



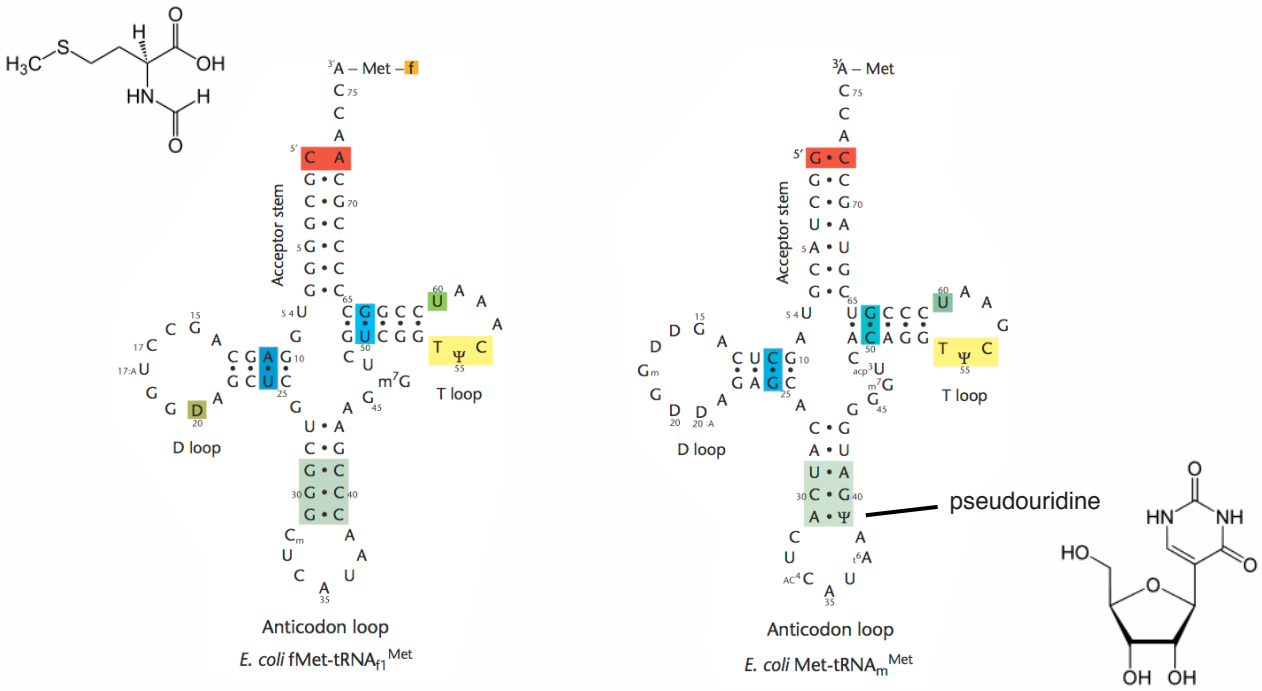
- We now turn to the mechanism of protein synthesis, a process called translation because the four-letter alphabet of nucleic acids is translated into the entirely different twenty-letter alphabet of proteins.
- Translation is a conceptually more complex process than either replication or transcription, both of which take place within the framework of a common base-pairing language. Befitting its position linking the nucleic acid and protein languages, the process of protein synthesis depends critically on both nucleic acid (somewhat more important) and protein factors (somewhat less important).
- Protein synthesis takes place in ribosomes — enormous complexes containing three large RNA molecules and more than 50 proteins.
- We will first take a look at the tRNA and the enzymes that ‘charge’ them with the correct amino acids.
- The sequence of amino acids in a protein is translated from the nucleotide sequence in mRNA. In which direction is the message read? The direction of translation is 5’ to 3’ in terms of the reading of the mRNA template. This corresponds to synthesis from the N-to-C terminus in terms of the protein product.
- The direction of translation has important consequences. Recall that transcription also occurs in the 5’-3’ direction. If the direction of translation were opposite that of transcription, only fully synthesized mRNA could be translated.
- In contrast, because the directions are the same, mRNA can be translated while it is being synthesized. In prokaryotes, almost no time is lost between transcription and translation. The 5’ end of mRNA interacts with ribosomes very soon after it is made, much before the 3’ end of the mRNA molecule is finished.
- An important feature of prokaryotic gene expression is that translation and transcription are closely coupled in space and time. Many ribosomes can be translating an mRNA molecule simultaneously. This parallel synthesis markedly increases the efficiency of mRNA translation.

Transfer RNA Molecules Have a Common Design



- The fidelity of protein synthesis requires the accurate recognition of three-base codons on messenger RNA. Recall that the genetic code relates each amino acid to a three-letter codon.
- A single amino acid could not recognize a codon by itself. Consequently, an amino acid is attached to a specific tRNA molecule that can recognize the codon by Watson-Crick base pairing. Transfer RNA serves as the adapter molecule that binds to a specific codon and brings with it an amino acid for incorporation into the polypeptide chain. Robert Holley first determined the base sequence of a tRNA molecule in 1965, as the culmination of 7 years of effort. Indeed, his study of yeast alanyl-tRNA provided the first complete sequence of any nucleic acid.
- The sequences of several other tRNA molecules were determined a short time later. Hundreds of sequences are now known. The striking finding is that all of them can be arranged in a cloverleaf-like secondary structure in which about half the residues are base-paired. Hence, tRNA molecules have many common structural features. This finding is not unexpected, because all tRNA molecules must be able to interact in nearly the same way with the ribosomes, mRNAs, and protein factors that participate in translation.
- A key feature of a tRNA molecule is the anti-codon that can base pair with a specific codon in the mRNA.

Initiator vs. elongator Met tRNA



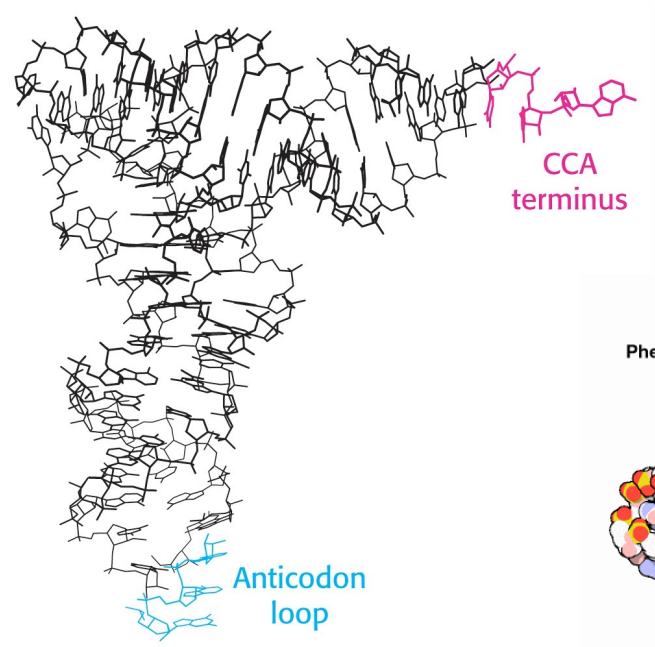
The two major differences are:
1) a missing base pair in the acceptor arm of the fMet-tRNA
2) 3 GC pairs in the anticodon stem of fMet-tRNA

http://en.wikipedia.org

Rasmussen, L. C., Laursen, B. S., Mortensen, K. K. and Sperling-Petersen, H. U. 2009. Initiator tRNAs in Bacteria and Eukaryotes. eLS.

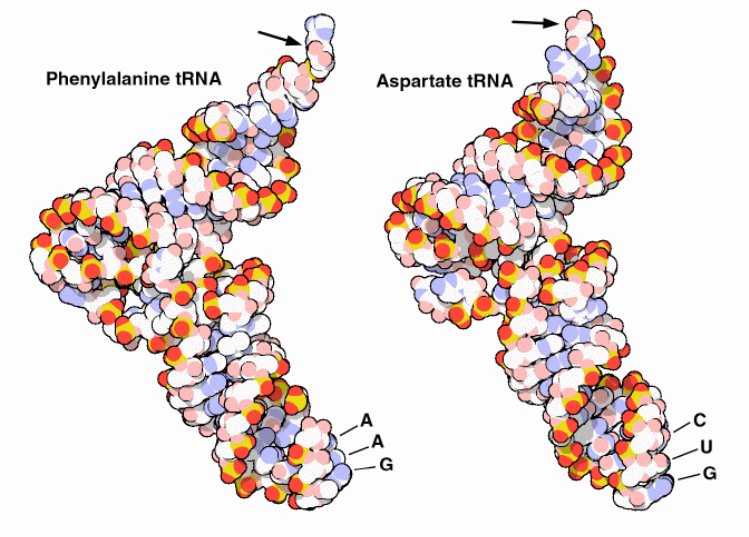
- Apparently, one tRNA synthetase is used to attach methionine to both tRNAs.
- The formyl group is added later by the enzyme methionyl-tRNA formyltransferase
- *Q: I just want to clarify primary, secondary, and tertiary structure. I was just thinking about tRNA, when it is in the three leaf clover shape would that be considered secondary structure? And then when you show it as the upside down L with double helices would that be considered the tertiary structure? I know that primary structure refers to the sequence but does that mean only linearly? Also when you show a protein as ribbons with alpha helices and b-strands would that be considered secondary or tertiary structure?*
- *A: Your explanation of primary, secondary, and tertiary structure for tRNA is correct. The primary structure is just the linear sequence of nucleotides. When we show a protein as any type of three dimensional structure, you are looking at the tertiary structure. To represent the secondary structure of a protein, you might imagine taking a 3D ribbon representation and pulling the ends to stretch it out into a line, but retaining the arrows that typically represent beta strands and the helices that represent alpha helix. Now you would have the secondary structure information but not the tertiary structure information. Alternatively, you might imagine secondary structure represented as a linear sequence where the amino acids are coloured blue for beta sheet residues and red for alpha helix.*

Transfer RNA molecules share a common three-dimensional structure



phenylalanine tRNA
(PDB entry 4tna)

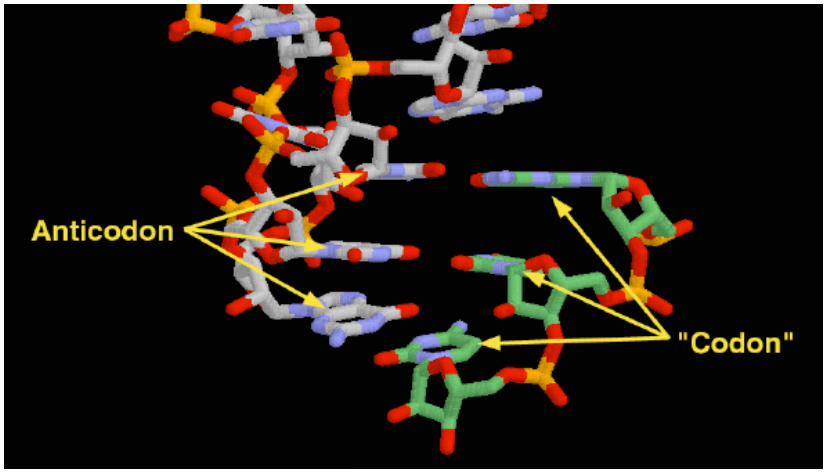
tRNAs for different amino acids all have very similar shapes



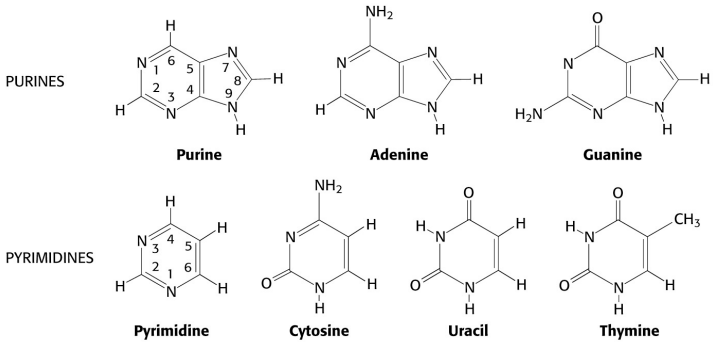
David S. Goodsell: The Molecule of the Month appearing at the PDB

- The three-dimensional structure of a tRNA molecule was first determined in 1974 through x-ray crystallographic studies carried out in the laboratories of Alexander Rich and Aaron Klug. The structure determined, that of yeast phenylalanyl-tRNA, is highly similar to all structures subsequently determined for other tRNA molecules.
- The most important properties of the tRNA structure are:
- The molecule is L-shaped (often represented in an orientation that looks like an upside down L).
- There are two apparently continuous segments of double helix. These segments are like A-form DNA, as expected for an RNA helix. The helix containing the 5' and 3' ends (pink in the right hand figure above) stacks on top of the helix that ends in the TψC loop (yellow) to form one arm of the L; the remaining two helices (blue and green) stack to form the other. Compare these colored regions with the schematic representation on the previous page.
- Most of the bases in the nonhelical regions participate in hydrogen-bonding interactions, even if the interactions are not like those in Watson-Crick base pairs.
- The CCA terminus containing the amino acid attachment site extends from one end of the L. This single-stranded region can change conformation during amino acid activation and protein synthesis.
- The anticodon loop is at the other end of the L, making accessible the three bases that make up the anticodon.
- An individual codon specifies a certain amino acid but there is no specific interaction between the codons and the amino acids themselves. Instead, the match is made by transfer RNA which translates the nucleotide language of codons into the amino acid language of proteins. This translation is physical and direct: at one end of each tRNA is an anticodon that recognizes the genetic code, and at the other end is the appropriate amino acid for that code.
- Two different molecules are shown here, phenylalanine tRNA (PDB entry 4tna) and aspartate tRNA (PDB entry 2tra). The two ends of the RNA chain are close to one another at the pointed end of the L-shaped structure, at the top in this illustration. The amino acid is added here, at the position shown by the arrows.

The anticodon on tRNA base pairs with a codon on mRNA



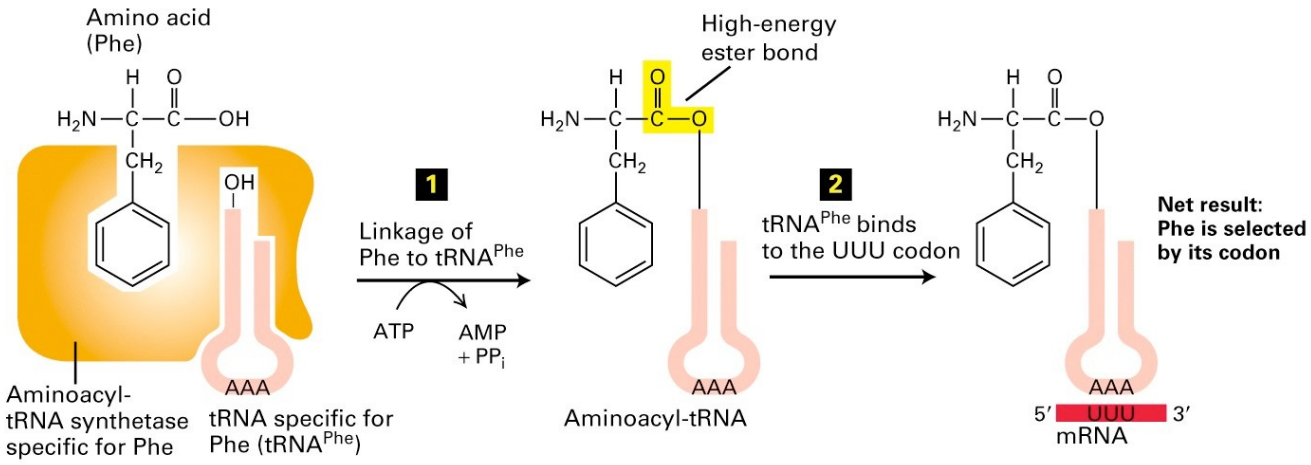
Can you identify all 3 bases of the anticodon?
Hint: there was no mRNA observed in this crystal structure and the 'codon' (notice the apostrophes) is in fact the anticodon of a 2nd copy of the tRNA.



David S. Goodsell: The Molecule of the Month appearing at the PDB

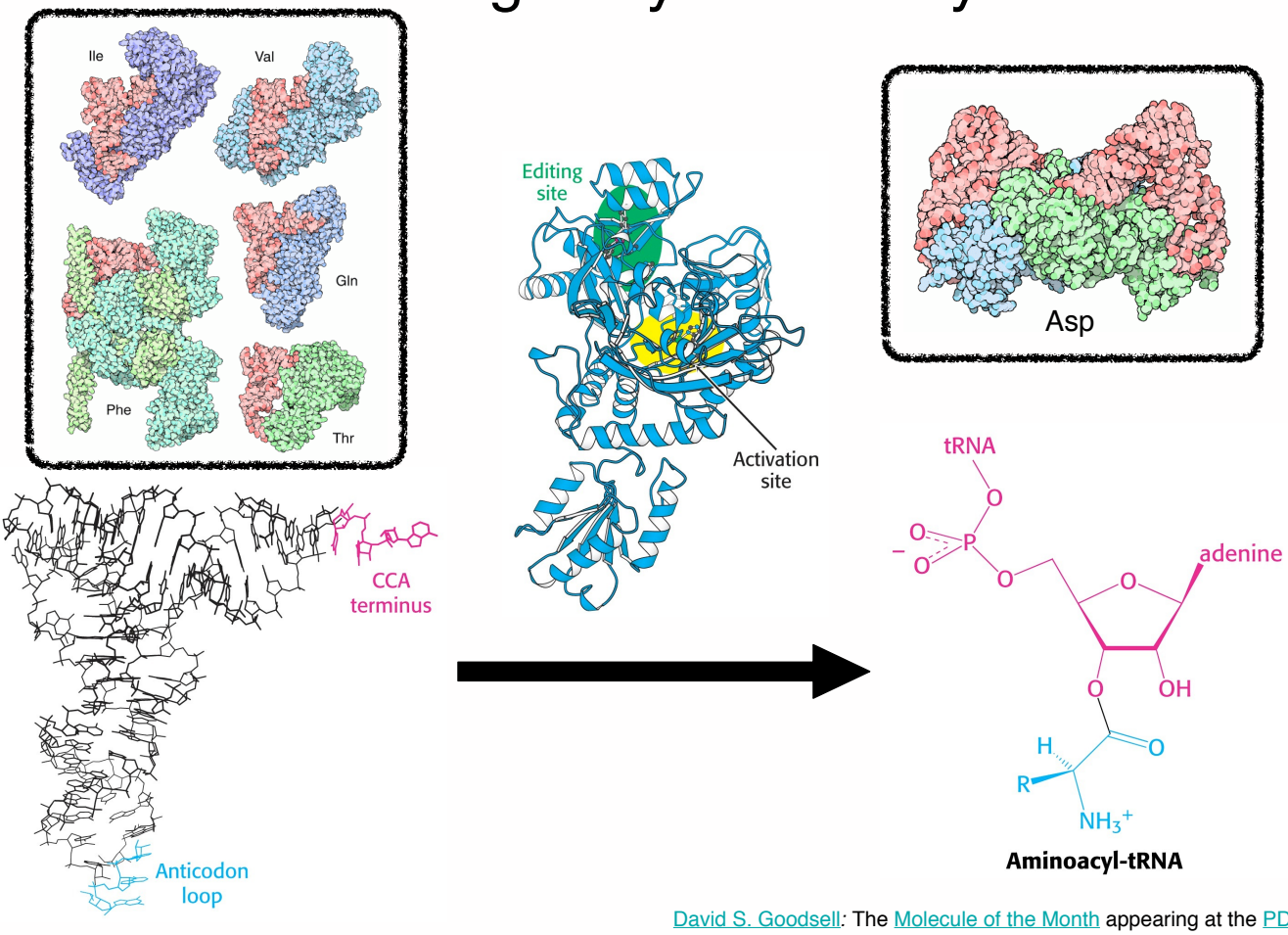
- Of course, the first thing we might want to look at when examining a tRNA structure is the anticodon. The structure 2tra, an aspartate tRNA, forms a dimer in the crystal lattice, with the anticodons of two separate molecules bound together. This is shown in the upper figure. The segment from the second molecule (with carbon in green in the illustration) will give you an idea of how the codon of an RNA message would bind to the tRNA anticodon.

tRNAs must first be connected to the correct amino acid



- Each tRNA is matched with its amino acid long before it reaches the ribosome. The match is made by a collection of remarkable enzymes, the aminoacyl-tRNA synthetases. These enzymes charge each tRNA with the proper amino acid, thus allowing each tRNA to make the proper translation from the genetic code of DNA into the amino acid code of proteins. Recognition of the codon specifying a given amino acid by a particular tRNA is actually the second step in decoding the genetic message.
- The first step, attachment of the appropriate amino acid to a tRNA, is catalyzed by a specific aminoacyl-tRNA synthetase. Each of the 20 different synthetases recognizes one amino acid and all its compatible, or cognate, tRNAs. These coupling enzymes link an amino acid to the free 2’ or 3’ hydroxyl of the adenosine at the 3’ terminus of tRNA molecules by a two-step ATP-requiring reaction.
- This enormously important step is the point at which “translation” takes place—at which the correlation between the amino acid and the nucleic acid worlds is made. In a sense, aminoacyl-tRNA synthetases are the only molecules in biology that “know” the genetic code. Their precise recognition of tRNAs is as important for high-fidelity protein synthesis as is the accurate selection of amino acids.
- Each tRNA molecule is recognized by a specific aminoacyl-tRNA synthetase. The ability of aminoacyl-tRNA synthetases to recognize their correct cognate tRNAs is just as important to the accurate translation of the genetic code as codon-anticodon pairing. Once a tRNA is loaded with an amino acid, codon-anticodon pairing directs the tRNA into the proper ribosome site; if the wrong amino acid is attached to the tRNA, an error in protein synthesis results.

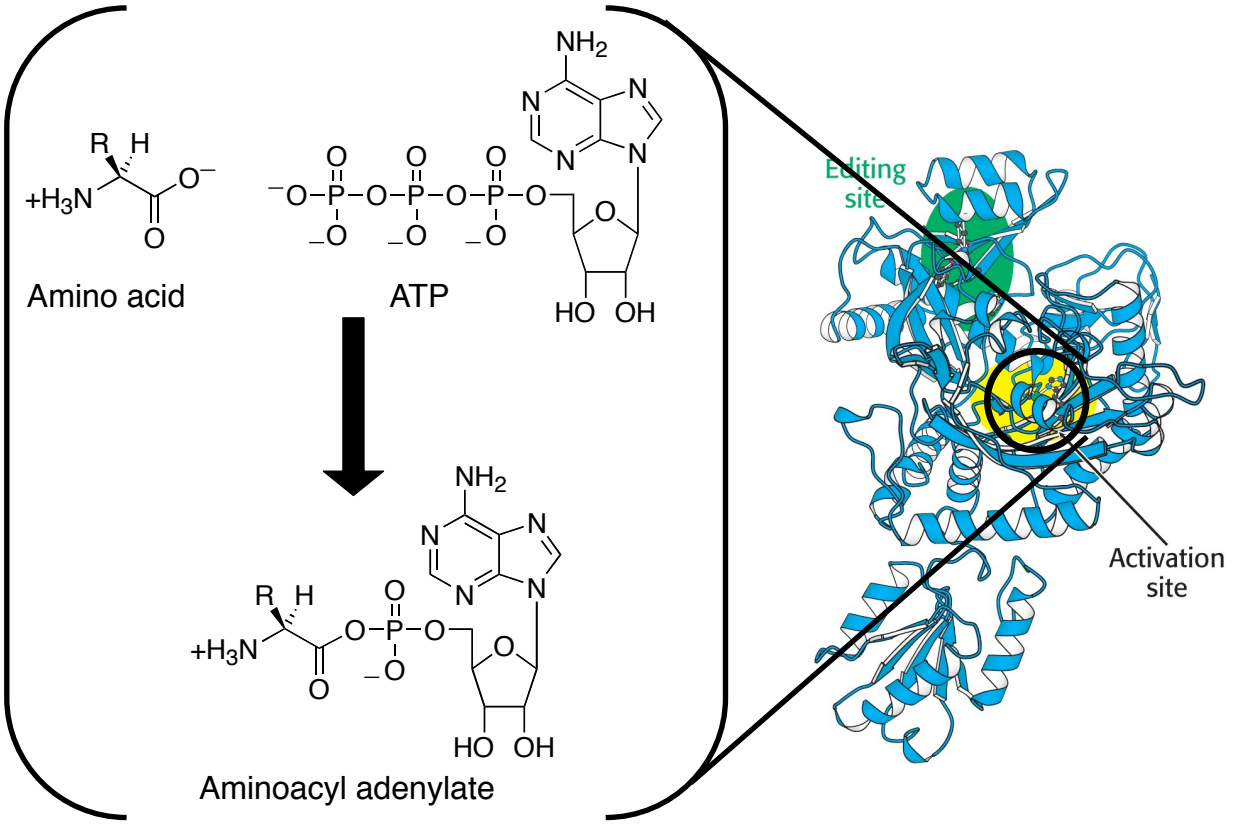
tRNAs are charged by a tRNA-synthetase



David S. Goodsell: The Molecule of the Month appearing at the PDB

- Most cells make twenty different aminoacyl-tRNA synthetases, one for each type of amino acid. These twenty enzymes are widely different, each optimized for function with its own particular amino acid and the set of tRNA molecules appropriate to that amino acid. The structures of nearly all of these different enzymes are available in the PDB.
- As you might expect, many of these enzymes recognize their tRNA molecules using the anticodon.
- Six complexes of an aminoacyl-tRNA synthetase with tRNA are shown. The isoleucine (entry 1ffy), valine (entry 1gax) and glutamine (entry 1euq) enzymes cradle the tRNA, gripping the anticodon loop (at the bottom in each tRNA), and placing the amino-acid acceptor end of the tRNA in the active site (at the top right in each tRNA). These all share a similar protein framework, known as "Type I," approaching the tRNA similarly and adding the amino acid to the last 2' hydroxyl group in the tRNA.
- The phenlyalanine (entry 1eiy) and threonine (entry 1qf6) enzymes are part of a second class of enzymes, known as "Type II." They approach the tRNA from the other side, and add the amino acid to the 3' hydroxyl on the last tRNA base.
- About half the aminoacyl-tRNA synthetases transfer the aminoacyl group to the 2' hydroxyl of the terminal adenosine (class I), and about half to the 3' hydroxyl (class II). In this reaction, the amino acid is linked to the tRNA by a high-energy ester bond and thus is said to be activated. The energy of this bond subsequently drives the formation of peptide bonds between adjacent amino acids in a growing polypeptide chain.
- An amino acid ester of tRNA is called an aminoacyl-tRNA or sometimes a charged tRNA.
- *Q: Just wondering how could the tRNA recognize the specific amino acid? I understand that the tRNA synth erase can recognize the anticodon on the tRNA, it's not a big deal. But how could it make the specific amino acid bind to its tRNA? I don't see any special structure in this amino acids.*
- *A: tRNAs do not recognize their amino acids. Once an amino acid has been attached to a tRNA, the tRNA will deliver it to the ribosome even if it is the wrong amino acid. tRNA synthetase enzymes are the enzymes that are responsible for attaching the correct amino acid to the correct tRNA. To do this, they must have an active site that recognizes the specific amino acid and they must also be able to bind specifically to the correct tRNA. The structures on p. 59 of the handouts shows how a tRNA synthetase is able to recognize the correct tRNA by binding to, at least, the anticodon part of the tRNA. The threonyl-tRNA synthetase example shows a close up of the amino acid binding site and shows how one particular tRNA synthetase is able to specifically bind the correct amino acid.*

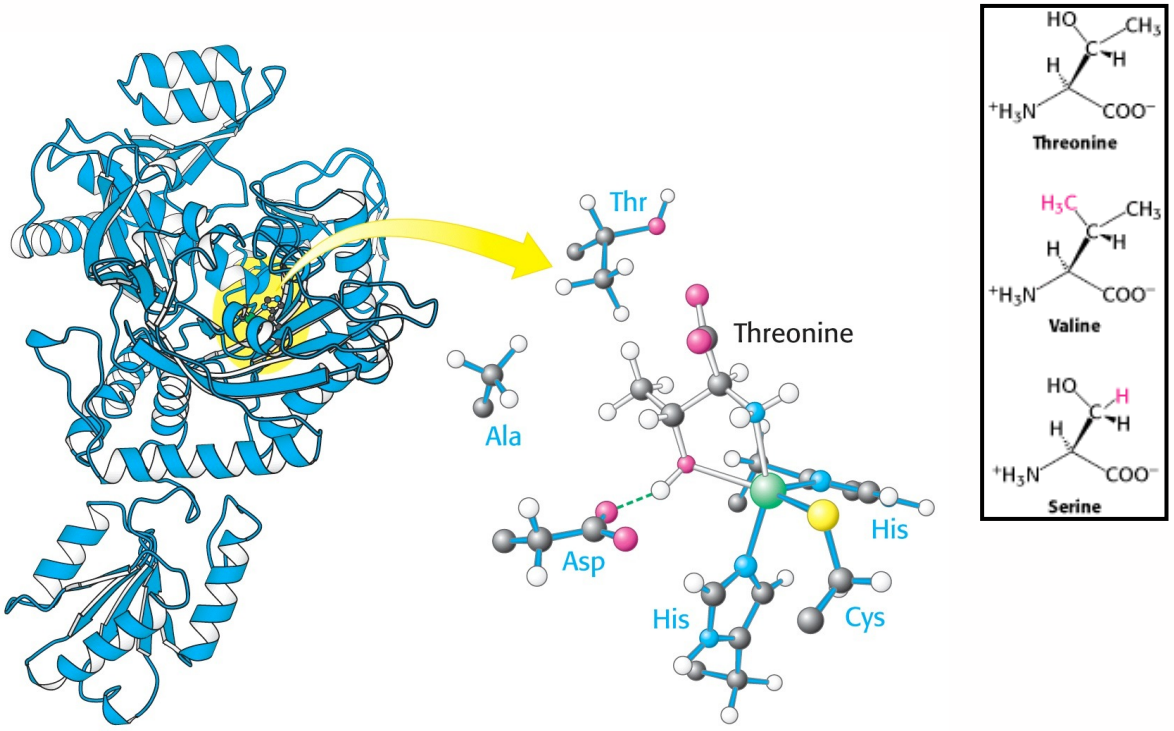
Amino Acids Are First Activated by Adenylation



- The chemistry for attachment of an amino acid to the tRNA involves two major steps of chemistry that happen in the same active site
- The first step is the formation of an aminoacyl adenylate from an amino acid and ATP.
- This activated species is a mixed anhydride in which the carboxyl group of the amino acid is linked to the phosphoryl group of AMP; hence, it is also known as aminoacyl-AMP.
- The activation and transfer steps for a particular amino acid are catalyzed by the same aminoacyl-tRNA synthetase. Indeed, the aminoacyl-AMP intermediate does not dissociate from the synthetase. Rather, it is tightly bound to the active site of the enzyme by noncovalent interactions. Aminoacyl-AMP is normally a transient intermediate in the synthesis of aminoacyl-tRNA, but it is relatively stable and readily isolated if tRNA is absent from the reaction mixture.
- The second step of the reaction is, of course, the transfer of the amino acid to the tRNA to form the aminoacyl tRNA. The 2' or 3' hydroxyl of the terminal adenosine of the tRNA would participate in a nucleophilic attack on the carbonyl of the aminoacyl-AMP intermediate. AMP is the leaving group, and amino acid remains attached to the tRNA.

tRNA-synthetases must bind amino-acids with very high specificity

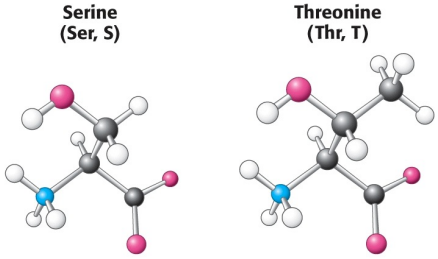
How does threonyl-tRNA synthetase discriminate against binding valine?
How does threonyl-tRNA synthetase discriminate against serine?



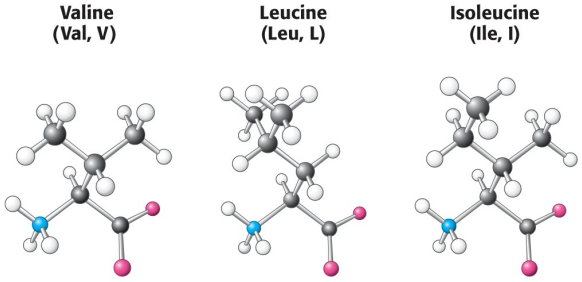
- Each aminoacyl-tRNA synthetase is highly specific for a given amino acid. Indeed, a synthetase will incorporate the incorrect amino acid only once in 10^4 or 10^5 catalytic reactions. How is this level of specificity achieved? Each aminoacyl-tRNA synthetase takes advantage of the properties of its amino acid substrate.
- Threonine is particularly similar to two other amino acids—namely, valine and serine. How can the threonyl-tRNA synthetase avoid coupling these incorrect amino acids to threonyl-tRNA? The enzyme contains a zinc ion in its active site - a feature that seems to be unique to threonyl-tRNA synthetase. Threonine coordinates to the zinc ion through its amino group and its side-chain hydroxyl group. The side-chain hydroxyl group is further recognized by an aspartate residue that hydrogen bonds to it. The methyl group present in valine in place of this hydroxyl group cannot participate in these interactions; it is excluded from this active site and, hence, does not become adenylated and transferred to threonyl-tRNA (abbreviated tRNAThr).
- The zinc site is less well suited to discrimination against serine because this amino acid does have a hydroxyl group that can bind to the zinc. Indeed, with only this mechanism available, threonyl-tRNA synthetase does mistakenly couple serine to threonyl-tRNA at a rate 10^-2 to 10^-3 times that for threonine. The enzyme needs a different strategy to prevent coupling with serine, and this is shown on the next slide.

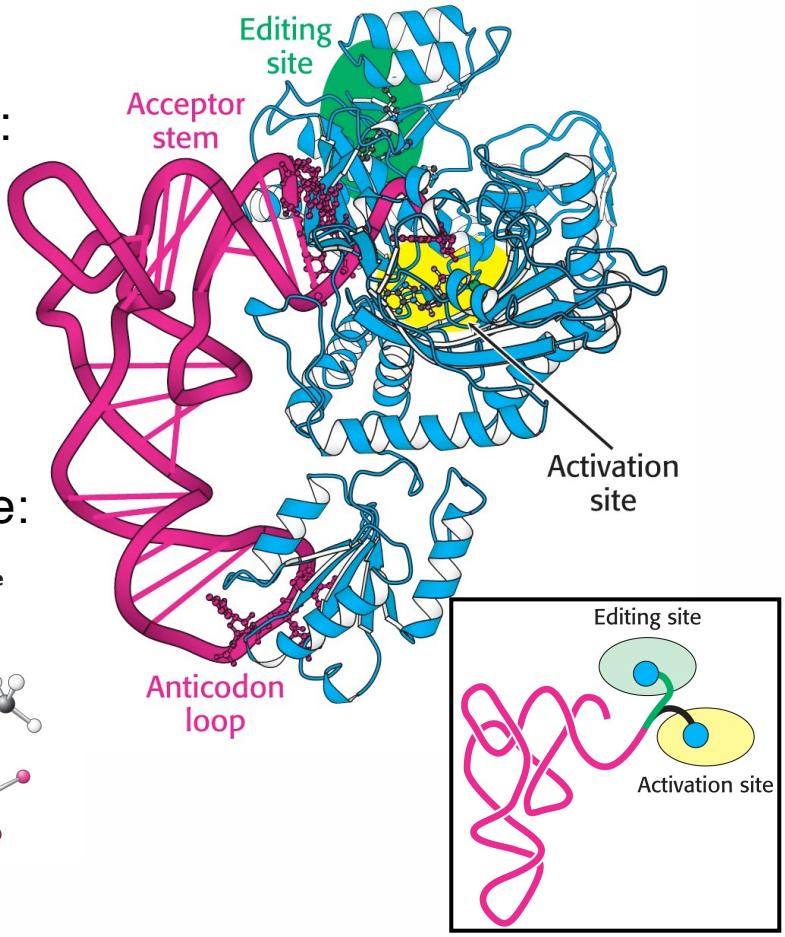
tRNA-synthetases can fix their own mistakes![68]

A very challenging case:



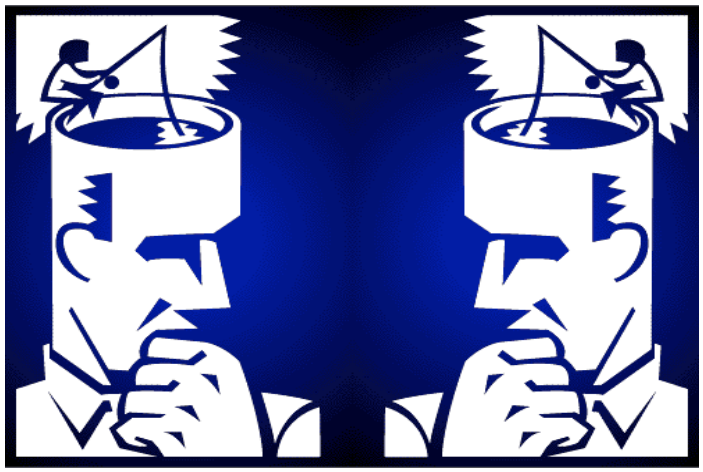
Another challenging case:





- Threonyl-tRNA synthetase can be incubated with tRNAThr that has been covalently (and artificially) linked with serine (Ser-tRNAThr); the tRNA has been "mischarged."
- Upon mixing there occurs a rapid hydrolysis of the aminoacyl-tRNA to form serine and free tRNA. In contrast, incubation with correctly charged Thr-tRNAThr results in no such hydrolysis reaction. Thus, threonyl-tRNA synthetase contains an additional functional site that hydrolyzes Ser-tRNAThr but not Thr-tRNAThr. This editing site provides an opportunity for the synthetase to correct its mistakes and improve its fidelity to less than one mistake in 10^4.
- The results of structural and mutagenesis studies revealed that the editing site is more than 20 Å from the activation site.
- This site readily accepts and cleaves Ser-tRNAThr but does not cleave Thr-tRNAThr. The discrimination of serine from threonine is relatively easy because threonine contains an extra methyl group; a site that conforms to the structure of serine will sterically exclude threonine. The opposite specificity (threonine preferred to serine) is much harder to achieve because serine is smaller than threonine and there are no additional steric clashes that occur when serine is placed in an active site that is intended for threonine.
- Isoleucyl tRNA synthetase faces a challenge that is similar to that faced by threonyl tRNA synthetase. The enzyme has an an isoleucine-shaped binding pocket in the active site of the enzyme. The slightly smaller amino acid valine, different by only a single methyl group, also fits nicely into this pocket, binding instead of isoleucine in about 1 in 150 times. This is far too many errors, so corrective steps must be taken. Like threonyl-tRNA synthetase, isoleucyl-tRNA synthetase has a second active site, which performs an editing reaction. Isoleucine does not fit into this site, but errant valine does. The mistake is then cleaved away, leaving the tRNA ready for a properly-placed isoleucine amino acid. This proofreading step improves the overall error rate to about 1 in 3,000.

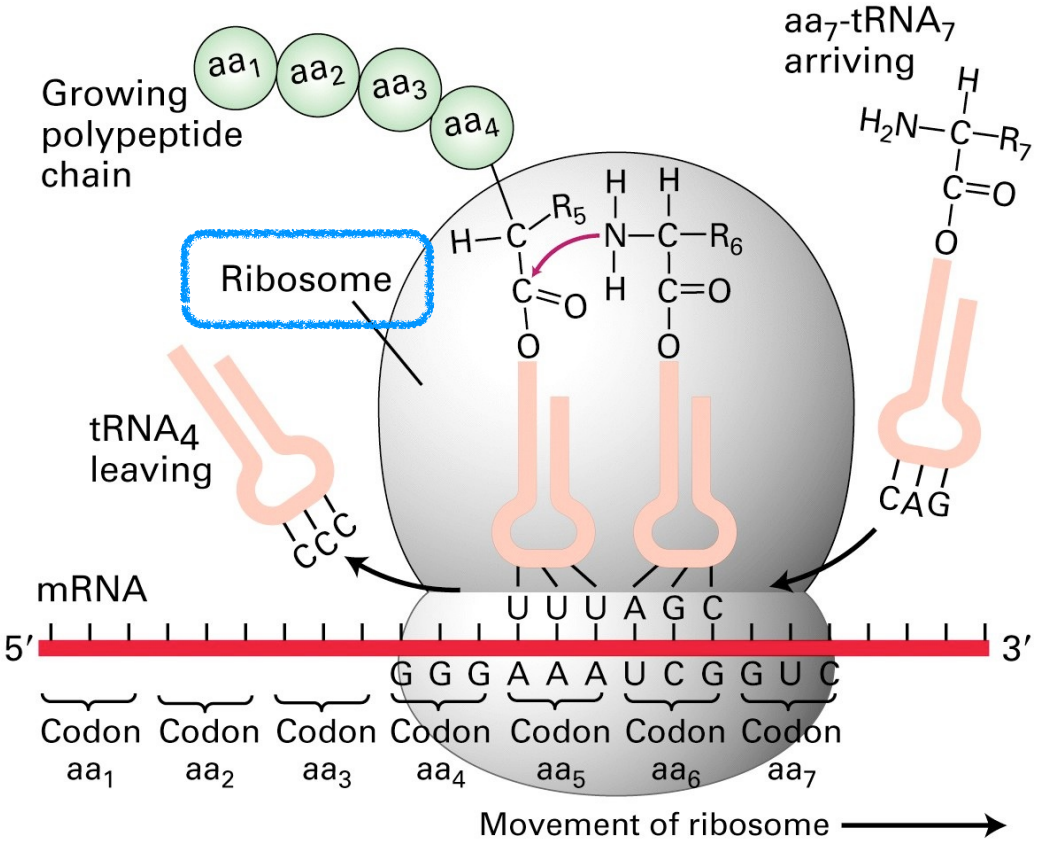
How could we artificially expand the genetic code?



- How would you switch one amino-acid for a very similar analogue (*i.e.*, isoleucine and trifluoromethyl isoleucine)?
- How might you add a 21st amino acid to the genome?
- How might you add the 22nd, 23rd, 24th….etc amino acid to the genome?

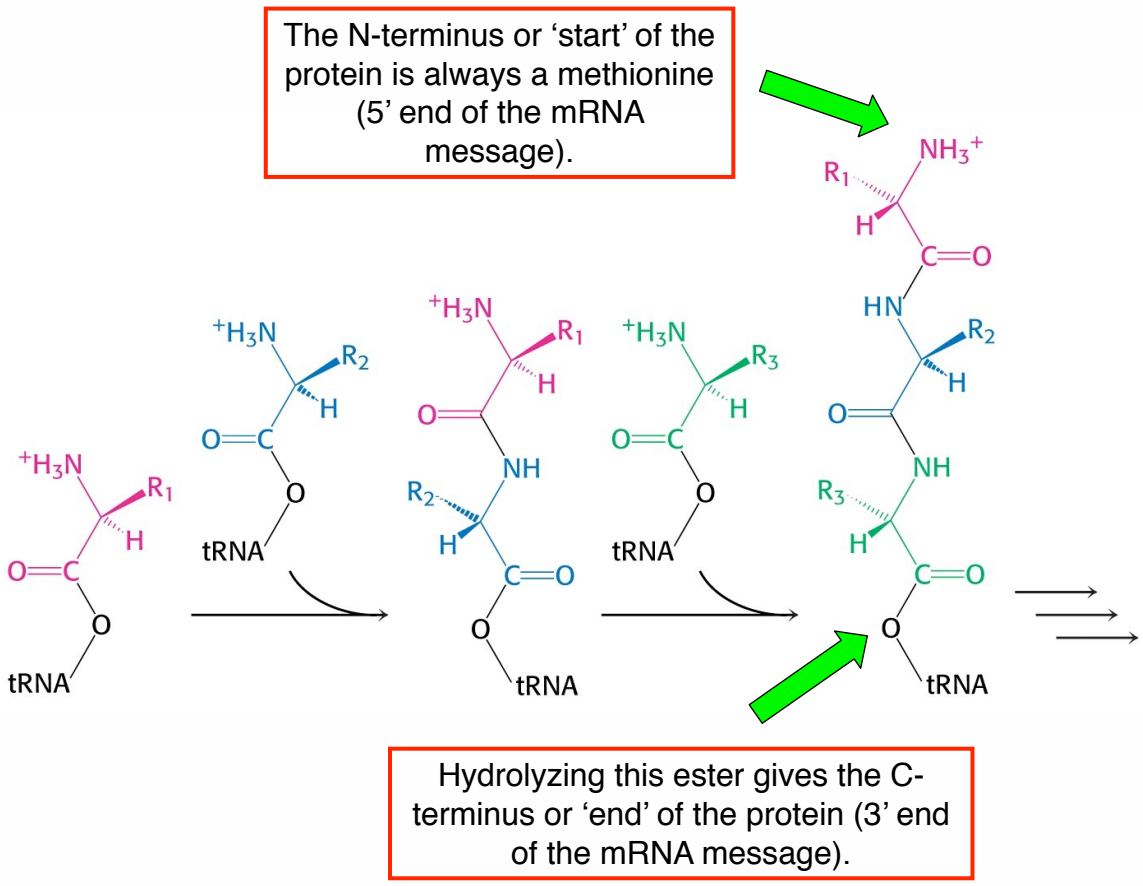
- If you want to read about some of the solutions that others have thought of, take a look at the homepage and publications of Peter Schultz at Scripps Research institute: http://schultz.scripps.edu/index.html

Overview of protein synthesis (translation)

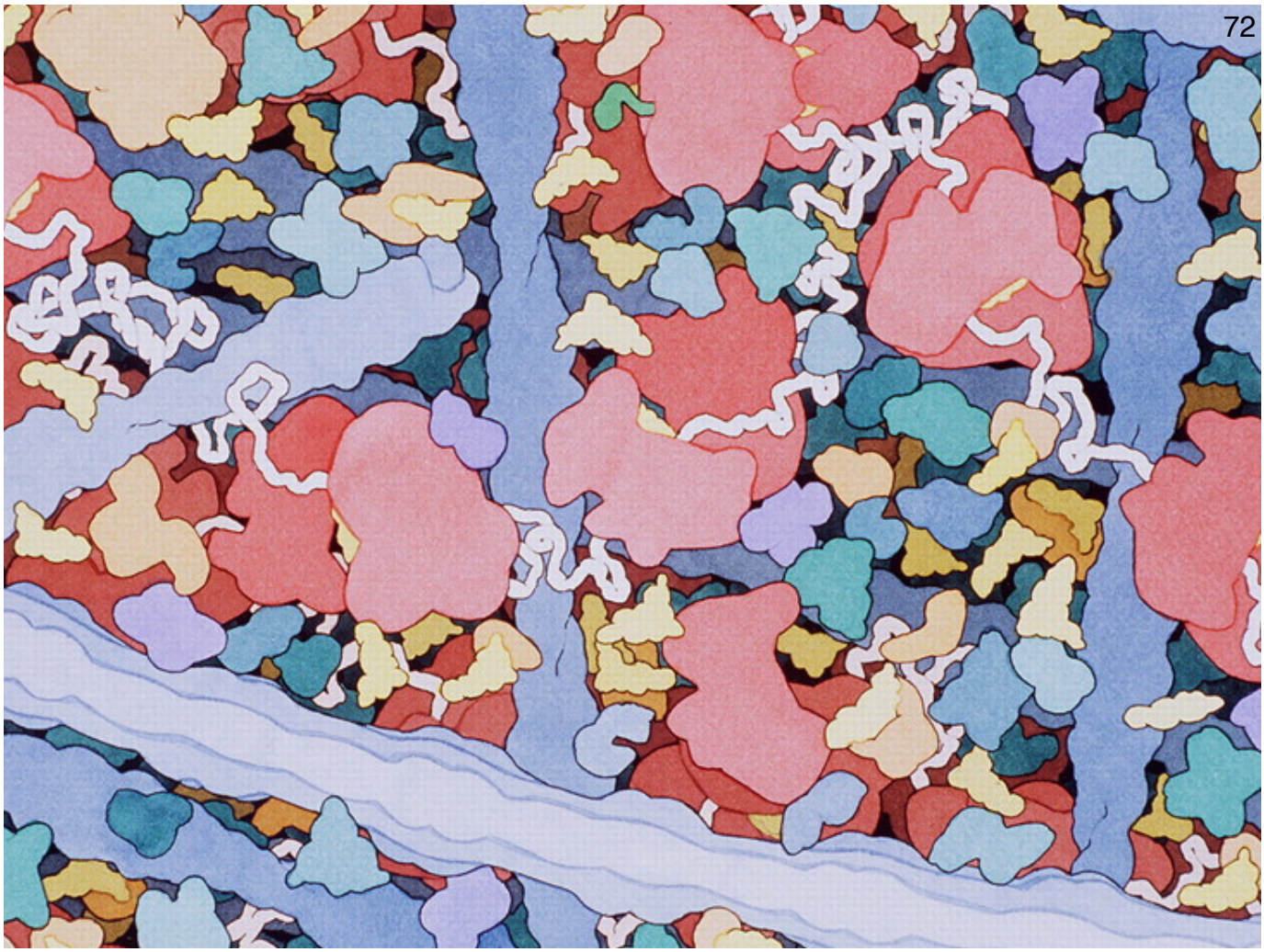


- We will now turn our attention to the ribosome itself.
- Ribosomes are the molecular machines that coordinate the interplay of charged tRNAs, mRNA, and accessory proteins that leads to protein synthesis.
- A ribosome is composed of several different ribosomal RNA (rRNA) molecules and more than 50 proteins, organized into a large subunit and a small subunit.
- The small ribosomal subunit contains a single rRNA molecule, referred to as small rRNA; the large subunit contains a molecule of large rRNA and one molecule each of two much smaller rRNAs in eukaryotes. The ribosomal subunits and the rRNA molecules are commonly designated in Svedbergs (S), a measure of the sedimentation rate of suspended particles centrifuged under standard conditions.
- The lengths of the rRNA molecules, the quantity of proteins in each subunit, and consequently the sizes of the subunits differ in prokaryotic and eukaryotic cells. (The small and large rRNAs are about 1500 and 3000 nucleotides long in bacteria and about 1800 and 5000 nucleotides long in humans.) Perhaps of more interest than these differences are the great structural and functional similarities among ribosomes from all species. This consistency is another reflection of the common evolutionary origin of the most basic constituents of living cells.
- Both the 30S and the 50S subunits can be reconstituted in vitro from their constituent proteins and RNA, as was first achieved by Masayasu Nomura in 1968. This reconstitution is an outstanding example of the principle that supramolecular complexes can form spontaneously from their macromolecular constituents.
- A ribosome can be dissociated into a large subunit (50S) and a small subunit (30S). These subunits can be further split into their constituent proteins and RNAs.

Polypeptide synthesis proceeds stepwise

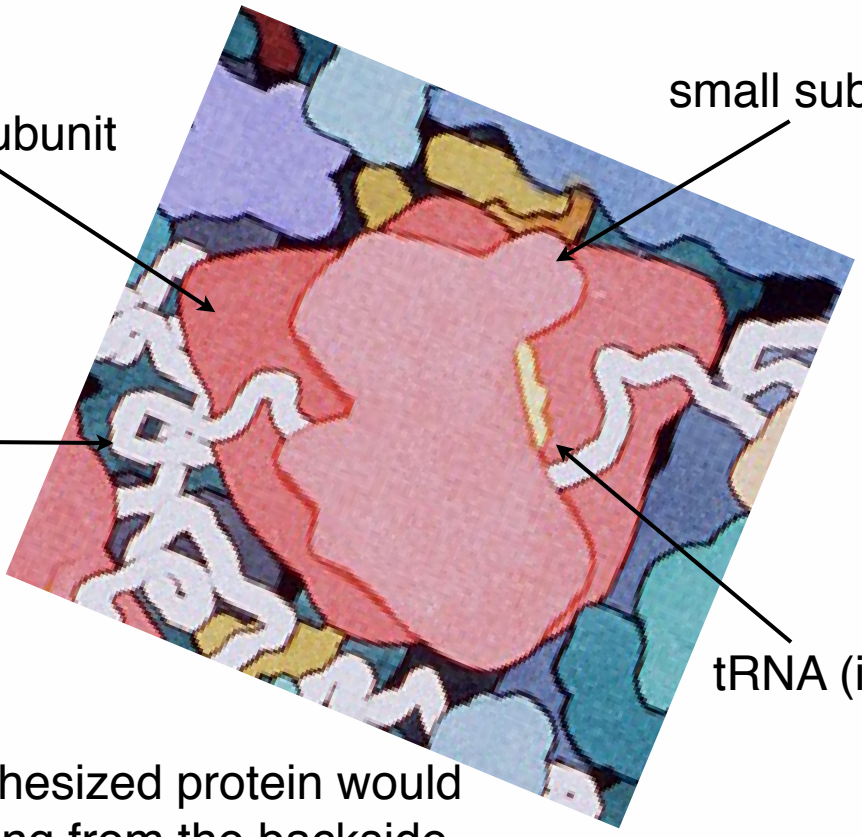


- The basics of protein synthesis are the same across all kingdoms of life, attesting to the fact that the protein-synthesis system arose very early in the evolution of life.
- A protein is synthesized in the amino-to-carboxyl direction by the sequential addition of amino acids to the carboxyl end of the growing peptide chain.
- As discussed in the previous section, the amino acids arrive at the growing chain in activated form as aminoacyl-tRNAs, created by joining the carboxyl group of an amino acid to a transfer RNA molecule.
- *Q: Say that the start codon is AUG, and the tRNA for Met would directly be ready and wait there for the translation? After the first second codon if we get a UUU, does everybody takes a look in there and the other tRNAs say to the Phe tRNA" hey, buddy, it's your turn, go ahead", and then it fits on the mRNA and give out the Phe it's carrying?*
- *A: It might seem inefficient, but the ribosome has to randomly bind to tRNA molecules (actually to complexes of tRNA and an accessory protein known as EF-Tu; http://www.rcsb.org/pdb/101/motm.do?momID=81) until it finds the right one. Only when the tRNA is a correct match for the codon, does EF-Tu release the tRNA and let it proceed with the peptidyl transfer reaction.*



- This picture represents the ribosome in its native habitat. The two subunits of the ribosome (shown in red) clamp around a snake-like messenger RNA (in white) and step down it, one codon at a time, building a new protein based on the encoded information. Scattered through the cytoplasm, many other molecules are needed (shown in yellow and orange), including transfer RNA and aminoacyl-tRNA synthetases, a collection of initiation, elongation and termination factors, and chaperones to help new proteins fold properly (David S. Goodsell: The Molecular Perspective appearing in The Oncologist)
- An *E. coli* ribosome is a ribonucleoprotein assembly with a mass of about 2700 kDa and a diameter of approximately 200 Å. The ~20,000 ribosomes in a bacterial cell constitute nearly a fourth of its mass.
- This is a great example of the best way to represent protein structures - trace their outlines!!

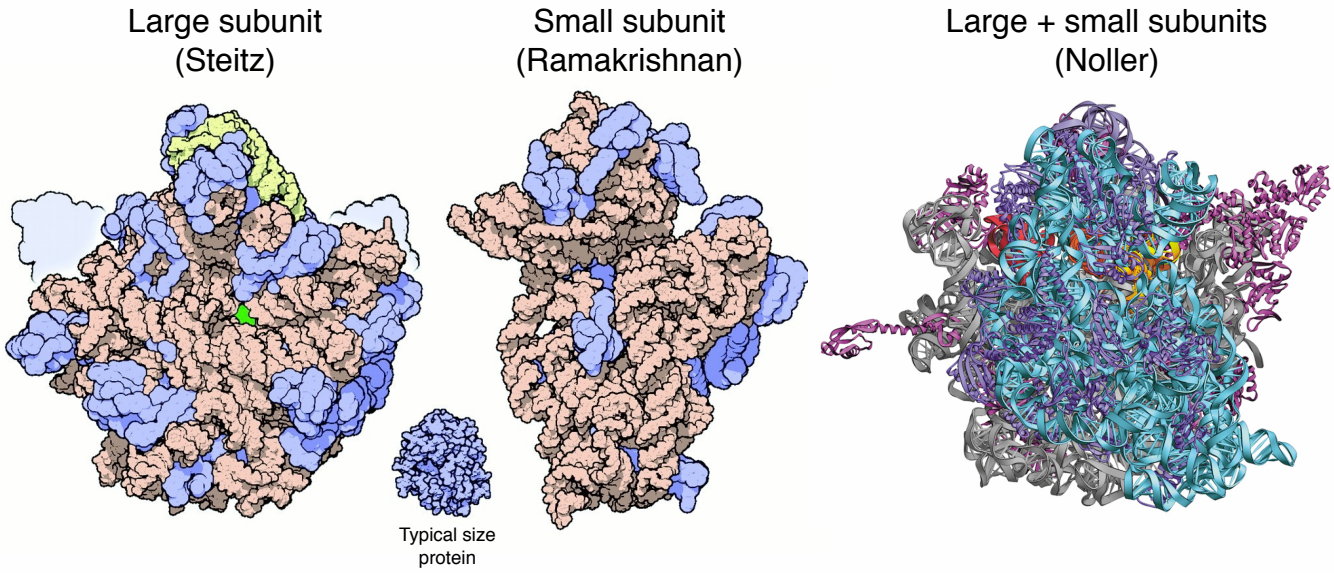
Overall architecture of the ribosome



newly synthesized protein would be emerging from the backside and is hidden from view.

- Contours of the bacterial ribosome and ribosomal subunits in two projections with designations of the main lobes. The small (30S) subunit is in lighter pink, the large (50S) subunit is in red. When synthesizing a new protein, the two subunits lock together with a messenger RNA trapped in the space between. The ribosome then walks down the messenger RNA three nucleotides at a time, building a new protein piece-by-piece.
- There is a spacious cavity or hole (entrance–exit channel) between the associated subunits where main functional sites of the ribosome, such as tRNA-binding sites, mRNA-retaining site of the small subunit and peptidyl transferase center of the large subunit, are localized.
- During translation, aminoacyl-tRNA molecules sequentially enter this intersubunit space from one side (entrance channel), deacylated tRNAs leave the ribosome from the other side (exit channel), and the chain of mRNA is drawn through the hole in the direction from its 5'-end to 3'-end. It is the passing of tRNA together with a paired cognate codon (nucleotide triplet) of mRNA through the intersubunit channel that is provided by the translocation mechanism of the ribosome.
- Electron microscopic studies of the ribosome at increasingly high resolution provided views of the overall structure and revealed the positions of tRNA-binding sites. Astounding progress on the structure of the ribosome has been made by x-ray crystallographic methods, after the pioneering work by Ada Yonath.
- http://www.weizmann.ac.il/sb/faculty_pages/Yonath/home.html

The Ribosome is very big, but it's structure has been determined at atomic resolution!



The Nobel Prize in Chemistry 2009

"for studies of the structure and function of the ribosome"

Venkatraman Ramakrishnan
Thomas A. Steitz
Ada E. Yonath

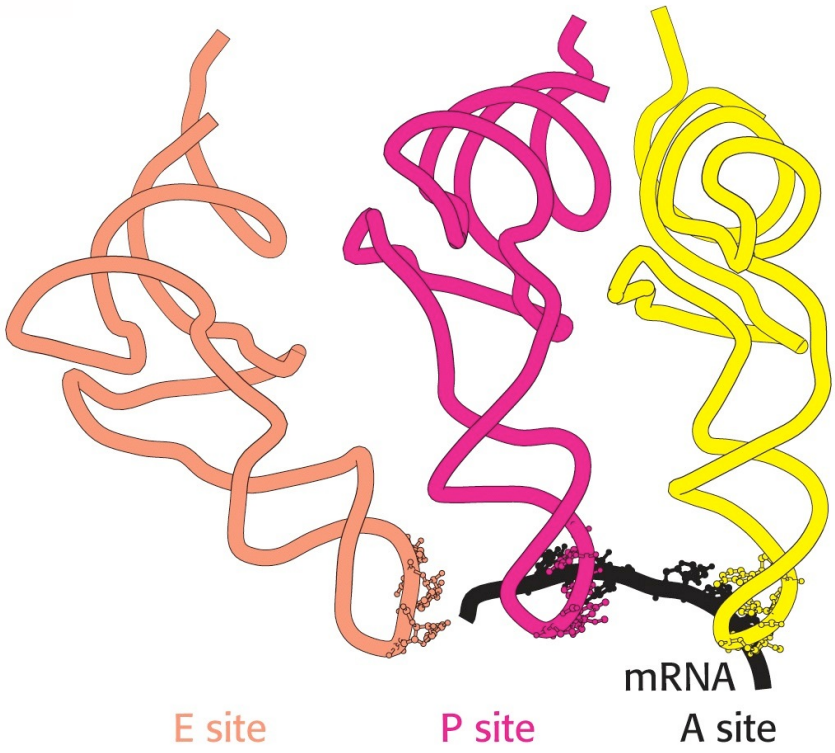
David S. Goodsell: The Molecular Perspective appearing in The Oncologist
http://rna.ucsc.edu/rnacenter/noller_lab.html

- The structures of both the 30S and the 50S subunits were determined at or close to atomic resolution in 2000, and the elucidation of the structure of intact 70S ribosomes was accomplished in 2001. The determination of this structure requires the positioning of more than 100,000 atoms. These structures provide an invaluable framework for examining the mechanism of protein synthesis.
- The x-ray crystal structure of the large subunit is shown on the left, and the x-ray crystal structure of the small subunit is shown in the middle. Of course, the term "small" is used in a relative sense here: both the large and the small subunits are huge compared to a typical protein (D-alanyl-D-alanine peptidase from entry 1cef). The putative catalytic adenine in the large subunit is shown in green. The structure of the complete 70S ribosome is shown on right.
    - Thomas A. Steitz from Yale University (http://www.yale.edu/steitz/)
    - Venki Ramakrishnan from the University of Utah (now at MRC in Cambridge, England) (http://www.mrc-lmb.cam.ac.uk/ribo/homepage/ramak/index.html)
    - Harry F. Noller at the University of California, Santa Cruz (http://rna.ucsc.edu/rnacenter/noller_lab.html)
- See also, the January 2010 Molecule of the Month by David Goodsell, http://dx.doi.org/10.2210/rcsb_pdb/mom_2010_1

There are 3 binding sites for tRNA in the ribosome

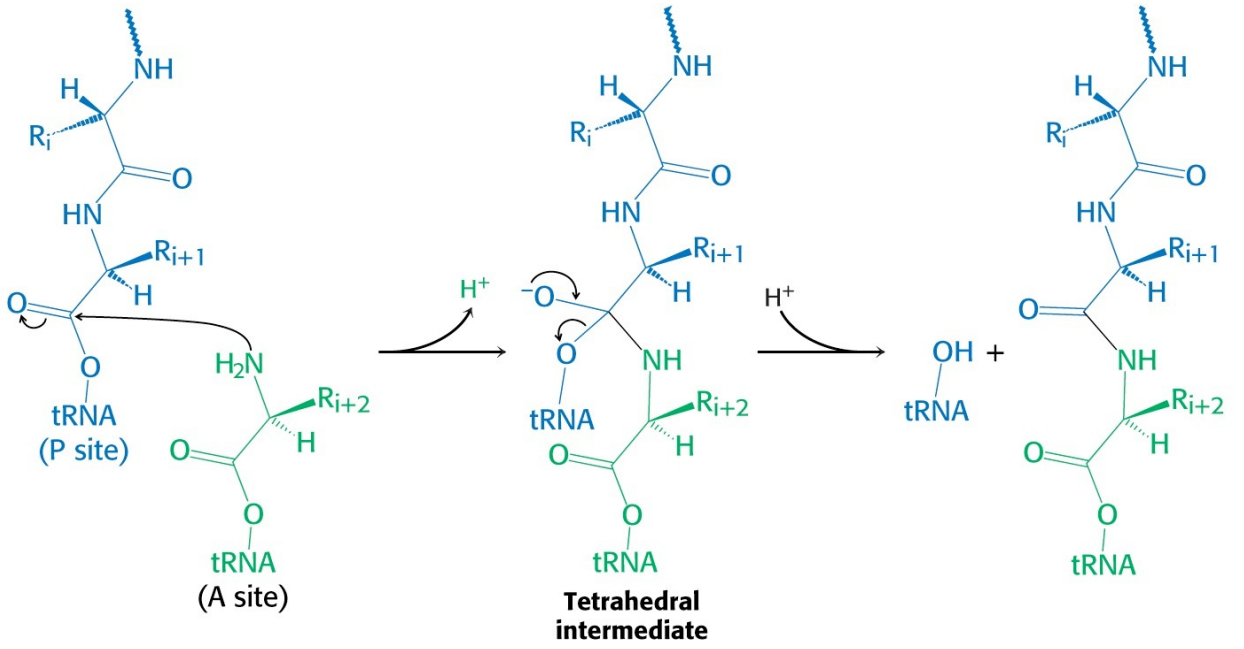
Here we are looking at a model of the tRNA and mRNA as they would be positioned inside the ribosome

A = aminoacyl
P = peptidyl
E = exit



- Originally, ribosomes were envisaged as having two binding sites for tRNA – the A (aminoacyl) site and the P (peptidyl) site. This was, first of all, the minimum number of binding sites that could account for the known mechanism of protein synthesis (attack of the amino group of aminoacyl-tRNA at the carbonyl group of peptidyl-tRNA), and secondly, it was consistent with virtually all of the available data at that time.
- The classical two-site model began to give way to the modern three-site model with the discovery of the E (exit) site by Nierhaus and co-workers. At least some of the early resistance to the E site was the lack of any theoretical reason to invoke a third site, and the tradition of favouring the simplest possible model (Why bother with an E site? Why not just discard the deacylated tRNA?). Eventually, the original findings were confirmed by several groups, and the paradigm expanded to include the E site.
- During a single step of protein synthesis a tRNA will progress through the ribosome in the following order
  - A (aminoacyl) site - when the tRNA is carrying its charged amino acid
  - P (peptidyl) site - when the tRNA is carrying the growing peptide chain
  - E (exit) site - last stop on the way out
- The most current theory and description of tRNA movement through the ribosome invoke ‘hybrid’ sites. On both the large and small subunit there is an A, P, and E site. For example: a tRNA that is fully in the A site is denoted A/A. However, immediately following peptide transfer, this tRNA is in a A/P hybrid state. The anticodon (on small subunit) is in the A site but the peptidyl end is in the P site. The anticodon end of the ribosome then translocates through the ribosome and the tRNA ends up in the P/P state.

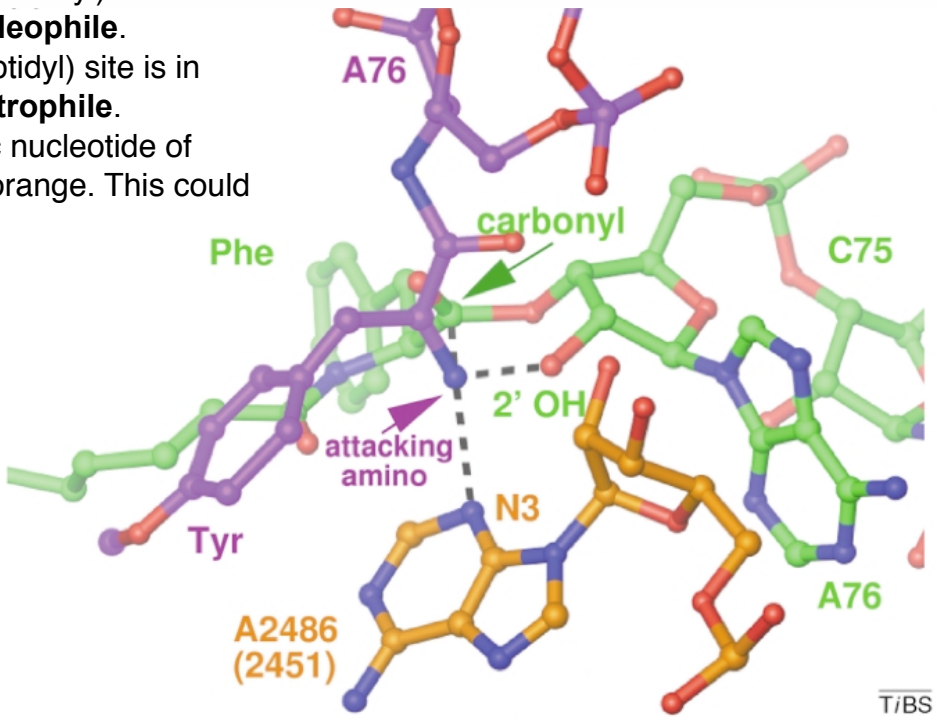
Mechanism of peptide bond formation



- The X-ray crystal structures of ribosomes have revealed that the amino group of the aminoacyl-tRNA in the A (aminoacyl) site is well positioned to attack the ester linkage between the tRNA in the P (peptidyl) site and the growing peptide chain.
- The peptidyl transferase center includes bases that promote this reaction by helping to form a nucleophilic -NH2 group on the A site aminoacyl-tRNA and by helping to stabilize the tetrahedral intermediate that forms.
- With the peptide bond formed, the peptide chain is now attached to the tRNA in the A site on the 30S subunit and the P site of the large subunit. The tRNA in the P site of the 30S subunit is now uncharged.
- Once this has happened, the ribosome can move onto the next codon in the mRNA. By shifting everything by one codon, the tRNA in the A site (with the peptide chain now attached it it) moves into the P site. The tRNA formerly in the P site moves to the E (exit) site. Note that the peptide chain stays more or less in the same place throughout this cycle, presumably growing into the tunnel.
- This cycle is repeated as the next required aminoacyl-tRNA move into the A site, allowing the polypeptide to be elongated indefinitely.

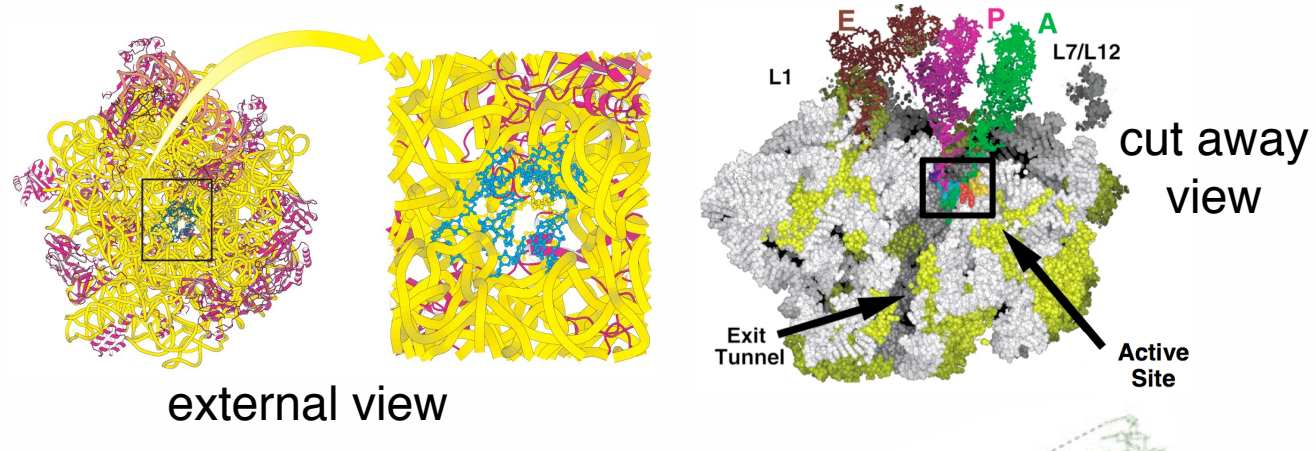
A close up (modeled) view of peptidyl transferase reaction

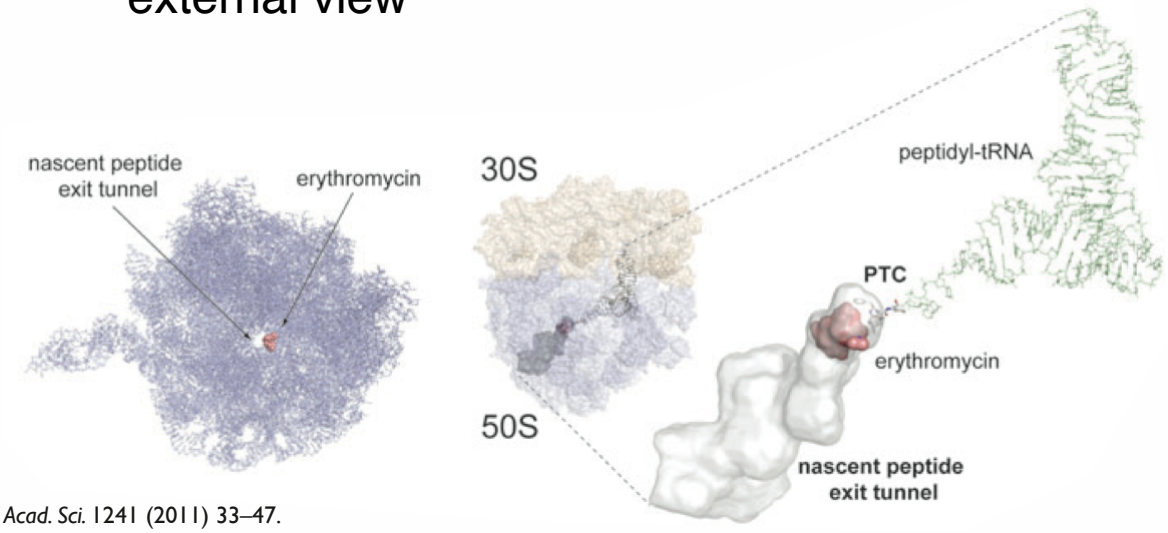
- The tRNA in the A (aminoacyl) site is in purple. This is the **nucleophile**.
- The tRNA in the P (peptidyl) site is in green. This is the **electrophile**.
- The proposed catalytic nucleotide of ribosome is shown in orange. This could be a **catalytic base**.



- A model of the peptidyl transferase centre of the large ribosomal subunit from H. marismortui with substrates bound to both the A-site and P-site. This model was obtained by superimposing the structure of an A-site substrate complex (PDB code: 1FGO) on the structure of a P-site substrate complex (PDB code: 1M90). The α-amino group of the A-site substrate (purple) is positioned for attack on the carbonyl carbon of the ester linking the peptide moiety of the P-site substrate (green). Possible hydrogen-bonding interactions involving the α-amino group and the N3 of A2486 (A2451 in E. coli) and the 2' OH of A76 are indicated. The 2' OH of A2486 (A2451 in E. coli) is also close enough so that is might interact.
  - J.L. Hansen et al., Structural insights into peptide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002), pp. 11670–11675.
  - P.B. Moore and T.A. Steitz, The structural basis of large ribosomal subunit function. *Annual Review of Biochemistry* 72 (2003) 813-850.

Ribosome New Protein exit tunnel

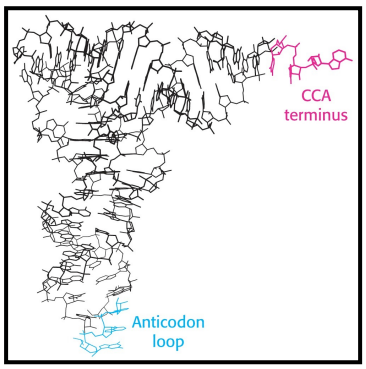




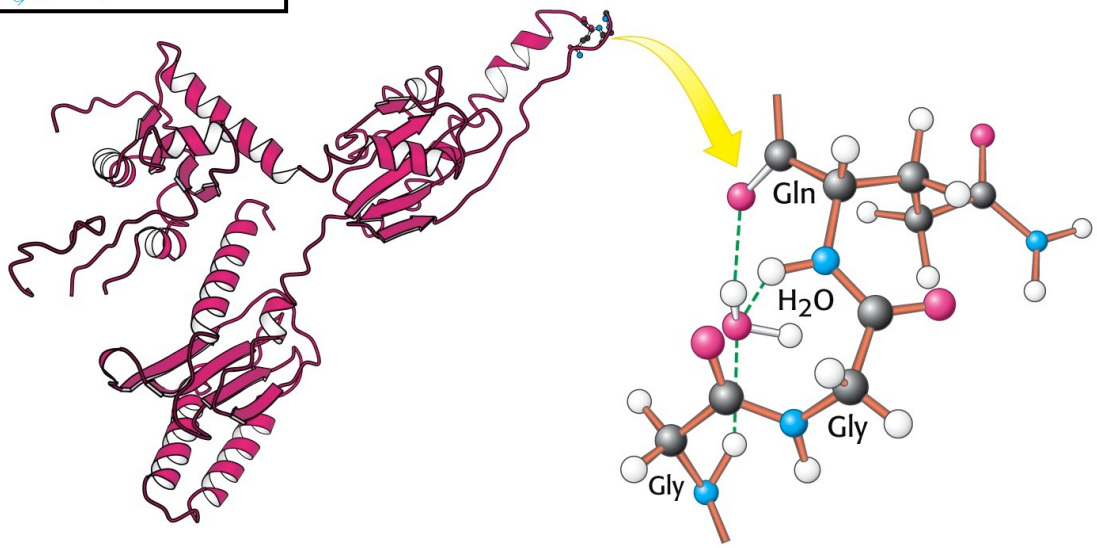
Ann. N.Y.Acad. Sci. 1241 (2011) 33–47.

- Peptide synthesis is happening right in the middle of this giant macromolecular complex so there isn't much room to hold the new protein.
- And it doesn't have to…. Because there is a backdoor!
- Shown in this figure is the tunnel exit on the backside of the ribosome.
- The tunnel appears very narrow but it seems big enough to fit an α-helix.
- It is not big enough to fit any sort of globular protein structure. Therefore the folding of proteins into three dimensional structures must occur after the protein has left the tunnel.
- The exit tunnel is target for some antibiotics - plugging the tunnel shuts down the synthesis of new proteins which would obviously be quite detrimental to the survival of a cell
- The most important class of such antibiotics are the macrolides, the most famous example of which is erythromycin A.
- These antibiotics bind to prokaryote ribosomes (i.e. bacteria) but not to eukaryote ribosomes (i.e. vertebrates).
- The tunnel seems to be quite conformationally mobile as opposed to a static tunnel structure.

Bringing protein synthesis to an end: release factors



Release factors work by a 'Trojan horse' strategy: disguised as a tRNA they sneak a water molecule into the active site of the ribosome - a site that usually excludes water molecules



- The codons UAA, UAG, and UGA designate chain termination. These codons are read not by tRNA molecules but rather by specific proteins called release factors. Binding of the release factors to the ribosomes releases the newly synthesized protein.
- One of the most impressive properties of the ribosome is not that it catalyzes peptide-bond formation; the formation of a peptide bond by the reaction between an amino group and an ester is a facile chemical reaction. Instead, a more impressive feature crucial to ribosome function is that the peptidyl-tRNA ester linkage is not broken by premature hydrolysis. The exclusion of water from the peptidyl transferase center is crucial in preventing such hydrolysis, which would lead to release of the polypeptide chain. The structure of a eukaryotic release factor reveals the strategy.
- The structure resembles that of a tRNA by molecular mimicry. The sequence Gly-Gly-Gln, present in both eukaryotes and prokaryotes, occurs at the end of the structure corresponding to the acceptor stem of a tRNA. This region binds a water molecule. Disguised as an aminoacyl-tRNA, the release factor may carry this water molecule into the peptidyl transferase center and, assisted by the catalytic apparatus of the ribosome, promote this water molecule's attack on the ester linkage, freeing the polypeptide chain.
- The detached polypeptide leaves the ribosome via the new protein exit tunnel mentioned on a previous slide.