Polymer biosynthesis

At the molecular level, the process that distinguishes living matter from non-living matter is the transition from monomers to polymers. Many crucial catalytic and structural components of cells are proteins. The genetic instructions for the synthesis of these proteins are found in polynucleotides. Proteins and polynucleotides are two of the major classes of cellular polymers. The details of the mechanisms of their syntheses are still active areas of research. A great deal is already known and the reader is referred to modern textbooks or the primary literature for details. Here, certain biophysical features will be highlighted with an emphasis on the gaps that still remain in our understanding.

Equilibrium thermodynamics for polymers

In order to develop a feel for the difficulty involved in generating significant concentrations of polymers, it is helpful to consider the volume per particle for different molarities. In table 1 the volume per particle is given for four representative molar concentrations.

Table 1

Molarity	Volume per particle
1 M	1.6 nm^3
1 mM	$1.6 \times (10\ nm)^3$
1 μM	$1.6 \times (100\ nm)^3$
1 nM	1.6 μ^3

Notice that at one nano-molar (nM), the volume per particle is 1.6 μm^3. This is roughly the volume of a coccus bacterium. Such a bacterium has a single molecule of DNA for its genome. Thus, its DNA concentration is roughly one nanomolar, nM. This may seem quite small, but for a large polymer it is actually an enormous value from an equilibrium point of view.

At standard state, a dipeptide costs ~ 1.5 kcal/mol to make from free amino acids. That is, $\Delta G^{0'} = +1.5 kcal/mol$ for

$$aa_1 + aa_2 \rightarrow aa_1 - aa_2 + H_2O$$

At 25 ^oC, the equilibrium constant for the spontaneous formation of dipeptide is given by

$$\frac{[aa_1 - aa_2][H_2O]}{[aa_1][aa_2]} = \exp\left[-\frac{\Delta G^{0'}}{RT}\right] = \exp\left[-\frac{1500}{592}\right] = 0.08$$

On the left-hand side, the concentration for water, $[H_2O]$, should be replaced by 1M since this is the convention for standard state even though the actual concentration is somewhat greater than 55.5 M. The true water concentration is built into the standard state Gibbs free energy, $\Delta G^{0'}$ as is signified by the superscript 0. Thus, the equilibrium concentration of dipeptide is given by

$$[aa_1 - aa_2] = 0.08 \times \frac{[aa_1][aa_2]}{1M}$$

If the amino acid concentrations are taken to be typical physiological values, i.e. about 1 mM, then $[aa_1\text{-}aa_2] = 8 \times 10^{-8}$ M. For a tripeptide, $\Delta G^{0'} = 2 \times 1.5 kcal/mol$ and $[aa_1\text{-}aa_2\text{-}aa_3] = 6 \times 10^{-12}$

M using the same assumptions for the amino acid concentrations. For a coccus bacterium volume, there would not be a single tripeptide present in equilibrium, much less a protein! Let that sink in !! The dehydration linkages in coenzymes and in polynucleotides are more costly and the coenzymes require typically four or five such bonds. Even the monomers for polynucleotide synthesis, the mononucleotide triphosphates, require four dehydration linkages, mostly at a high cost. Thus, in equilibrium, a coccus bacterium volume would contain no! coenzymes and no! mononucleotides, much less polynucleotides such as DNA or RNA! Making proteins and polynucleotides requires an elaborate molecular apparatus and the input and utilization of energy. The details of these processes occupy the following sections.

Gene directed protein biosynthesis

Genetic information is coded in sequences of nucleotide bases in DNA. The information storage form of DNA is a double helix of antiparallel strands of DNA joined together by a multitude of hydrogen bonds between complementary base pairs. The base pairing rules are A with T (or T with A) and G with C (or C with G). The A-T base pair has two hydrogen bonds and the G-C base pair has three. The covalent backbone of DNA is the phosphodiester chain of deoxyribose, phosphate, deoxyribose, phosphate, etc.. The phosphates join the deoxyribose units in 5' to 3' linkages (these are dehydration linkages). This structure confers a polarity on a DNA single strand. The DNA duplex is made from two antiparallel strands, forming a right-handed helix. During cell division, DNA is *replicated*. Replication requires a number of obligatory, large enzymes (the DNA polymerases have MW ~100kd). DNA cannot self-replicate; if it could, it would be the quintessential autocatalytic molecule. Even though DNA backbones are made of deoxyribose, phosphate, deoxyribose, phosphate etc. sequences, the precursors are nucleotide triphosphates, not nucleotide monophosphates. The extra phosphates serve to provide activation energy for polymerization. When DNA is hydrolyzed, the products are nucleotide monophosphates. These need to be recharged to triphosphates if synthesis of DNA from them is to occur. The source of the extra phosphates is ATP (*a*denosine*trip*hosphate). Thus, replication of DNA requires a battery of enzymes, nucleotide monophosphates and ATP for monomer activation.

DNA is also *transcribed* in order to direct protein synthesis. The direct product of a DNA gene is an mRNA (*m*essenger RNA) *transcript*. Recall that RNA differs from DNA in that it contains ribose rather than deoxyribose and it contains U (uracil) rather than T. The mRNA is *translated* into protein. The mRNA transcript of a gene is transcribed with a polarity antiparallel to the gene itself. All base pairing interactions between polynucleotides, whether DNA or RNA, have this antiparallel feature. Translation does not occur directly on the mRNA. The process of translation is the most complex molecular process inside the cell. The reading of the mRNA is coordinated by a *ribosome*, an enormous complex of proteins and rRNA (*r*ibosomal RNA). It also requires tRNA's (*t*ransfer RNA's) that have at one end an anti-codon that reads the codons on the mRNA's by antiparallel base pairing and at the other end carry the cognate amino acid. The DNA not only has genes for the mRNA's but also for the rRNA's and the tRNA's. While base pairing rules govern the reading of mRNA codons by tRNA anti-codons, the attachment of the cognate amino acid to the correct tRNA is governed by a class of pure protein enzymes called aaRS's (*a*mino*ac*yl-t*R*NA *s*ynthetases). How these enzymes identify the cognate tRNA's and amino acids is known as the *mystery of the "second" code (*or*: mystery of the "operational" code)*. It is still largely unsolved. As will be seen, this problem is the *mysterium tremendum*. In

order for protein synthesis to occur, the amino acids must be activated and the activation energy comes from ATP. In addition, the smooth functioning of the ribosomal mRNA complex requires energy from GTP as well.

A strand of DNA running in the 3' to 5' direction is transcribed into RNA running in the 5' to 3' direction. By convention, the codons are defined relative to the mRNA and run 5' to 3'. The corresponding anti-codons of the tRNA's run 3' to 5', in parallel with the DNA gene. The translated protein is also polarized with an N-terminus (amino end) and a C-terminus (carboxyl end). The mRNA polarity of 5' to 3' corresponds with the protein polarity of N-terminus to C-terminus.

Ribosomes are cellular organelles. They are not as large as the membranous organelles such as the mitochondria, the Golgi apparatus and the lysosomes, but they are just as complex. An E. Coli ribosome has a mass of ~ 2.5 x 10^6 d and a sedimentation coefficient of 70S (svedbergs). It is ~ 25 nm across in its largest dimension. It dissociates into two unequal subunits, the small (30S) and the large (50S). The small subunit has a 16S rRNA and 21 polypeptides, whereas the large subunit has a 5S rRNA and a 23S rRNA and 31 polypeptides. Up to 20,000 ribosomes are in a single E. Coli cell and account for ~ 80% of its RNA and 10% of its protein. The E. Coli 16S rRNA has 1542 nucleotides, the 5S rRNA has 120 and the 23S rRNA has 2904. Each of these polymeric constituents has been completely sequenced.

Ribosomal subunits are self-assembled. In the early 1970's, Masayasu Nomura determined how this occurs. There is a quite strict order to the assembly process. Thus, when the cell needs more ribosomes for protein synthesis, it produces the constituent proteins and rRNA's in a controlled order to facilitate self-assembly. Nomura was able to disassemble and reassemble ribosomes that were functional in *in vitro* assays.

The genetic code

Amino acids do not interact directly with mRNA. Adaptors, called tRNA's, enable the codon sequences on the mRNA to be translated into amino acid sequences. The genetic code is virtually the same in all organisms.

Genetic code

1rst base	2nd base: U	C	A	G	3rd base
U	phe	ser	tyr	cys	U
U	phe	ser	tyr	cys	C
U	leu	ser	stop	stop	A
U	leu	ser	stop	trp	G
C	leu	pro	his	arg	U
C	leu	pro	his	arg	C
C	leu	pro	gln	arg	A
C	leu	pro	gln	arg	G
A	ile	thr	asn	ser	U
A	ile	thr	asn	ser	C
A	ile	thr	lys	arg	A
A	met	thr	lys	arg	G
G	val	ala	asp	gly	U

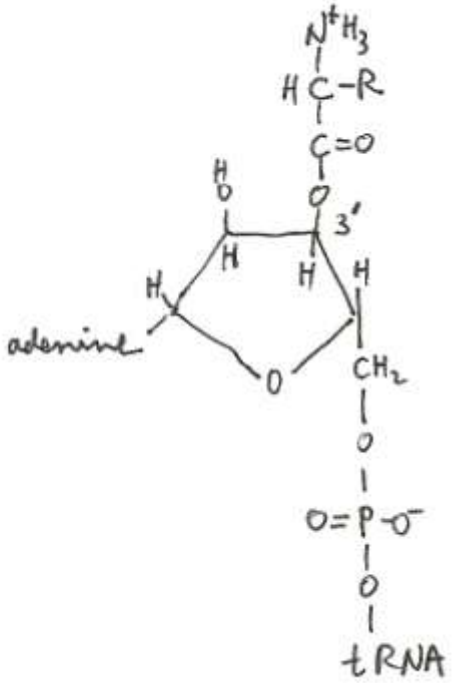
G	val	ala	asp	gly	C
G	val	ala	glu	gly	A
G	val	ala	glu	gly	G

An amino acid is coded for by a triplet of bases. With 4 bases this makes 64 possibilities. The code is highly degenerate it that 20 amino acids are coded for by 61 triplets, with three triplets reserved for *punctuation*, i.e. either initiation or termination signals. Much of the degeneracy can be traced to the codon base at the 3' end. As few as one codon per amino acid to as many as 6 are found. A fundamental question is whether this code is a frozen accident for life here on Earth, or whether it has a physico-chemical origin, making it universal.

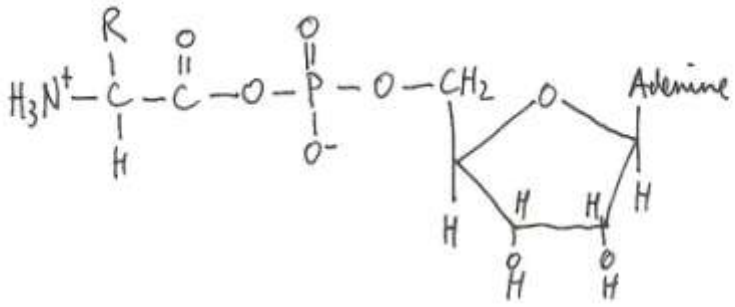
The tRNA's contain anti-codons at one end that read the codons by anti-parallel base pairing. At the other end of the tRNA is the amino acid attachment site on the tRNA 3' terminus. The aaRS's catalyze the attachment of the amino acids to the tRNA's at the expense of ATP.

$$tRNA + aa + ATP \rightarrow aa - tRNA + AMP + PP$$

in which AMP denotes adenine monophosphate and PP denotes pyrophosphate. The linkage between the carboxyl group of the amino acid and the 3' -OH group of the terminal ribose of the tRNA is a high energy ester without either P or S. This is uniquely unusual.



The intermediate is the aminoacyl adenylate, aa-AMP



The aaRS recognizes and binds the amino acid and the cognate tRNA, that for amino acid aa is labelled $tRNA^{aa}$. It also binds ATP and forms the intermediates aa-AMP and PP. The PP is released and subsequently hydrolysed, and the aa-AMP reacts with the bound $tRNA^{aa}$ to form $aa\text{-}tRNA^{aa}$. All of this occurs enzyme bound on the aaRS, but free aminoacyl adenylates can be prepared in the laboratory as free chemical species.

There is at least one aaRS for each of the 20 amino acids. The structural similarity of all tRNAs suggests that the aaRS's should all be closely related since they perform the same tasks. This is not so, they are a very diverse group of enzymes. Four types of subunit structure are found: α, α_2, α_4 and $\alpha_2\beta_2$. This means that some have just one subunit of type α, the α's. Others have multiple subunits of type α, the α_2's and α_4's. Some have two copies of two types of subunits, the $\alpha_2\beta_2$'s. For different amino acids, the subunits involved are of different sequence and structure, even though each is referred to as, say, α. Subunit sizes range from 334 to 1112 amino acids residues. Little sequence similarity exists from one aaRS species to another. Two major classes of aaRS's are found: class I and class II. Each has the same 10 members in all organisms. Class I aaRS's require anti-codon recognition to aminoacylate their cognate tRNA's. Class II aaRS's do not interact with the anti-codons. Class I aaRS's aminoacylate their tRNA's 3' terminal 2'-OH group whereas class II aaRS's charge the 3' terminal 3'-OH group instead.

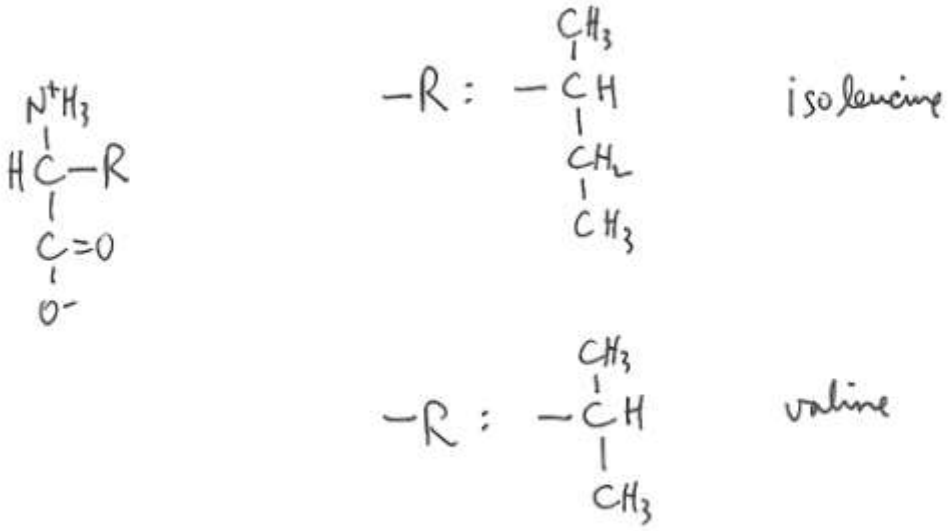
Class I	Class II	p+ ; polar +
glu p-	asp p-	p- ; polar -
arg p+	his p+	pn ; polar neutral
cys pn	lys p+	np ; non-polar

met pn	ser pn	
tyr pn	thr pn	
gln np	asn np	
ile np	gly np	
leu np	ala np	
val np	pro np	
trp np	phe np	
lys* p+		

The mystery of the second (operational) code

In order for the base pairing code to work, it is essential that cognate aa's are attached to their cognate $tRNA^{aa}$'s. This is the task of the aaRS's. It is clear that how they do this has an evolutionary history, given the diversity of aaRS structure and mechanisms. How they do their job is *the mystery of the second (operational) code*. It is probable that amino acid recognition is similar to substrate recognition by any enzyme. Recognition of the tRNA's is the real problem. Half of the aaRS's, those in class II, don't even need the anti-codon of the tRNA to be present to recognize the correct tRNA. Much work is still going on to explain this recognition process.

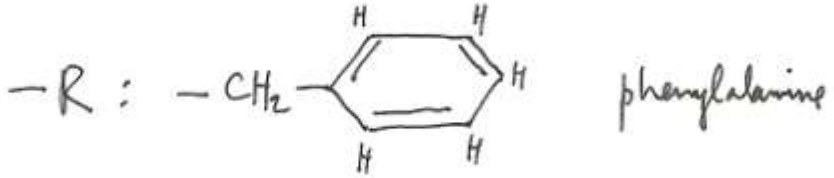
Proofreading enhances fidelity of amino acid attachment to tRNA. For example, IleRS (isoleucine aaRS) transfers 50,000 isoleucines to $tRNA^{Ile}$ for every valine it transfers incorrectly. These amino acids have similar residues.

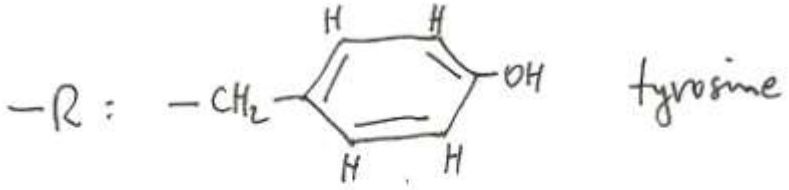


The isoleucine binding site of IleRS is large enough to bind isoleucine but not larger aa's. However, valine only lacks a single methyl group relative to isoleucine and can easily fit into the isoleucine binding site. The binding free energy of the $-CH_3$ group is estimated to be $\Delta G \sim -2.87$ kcal/mol. Thus the preference for isoleucine over valine is only

$$\exp\left[-\frac{\Delta G}{RT}\right] = \exp\left[\frac{2870}{592}\right] = 127$$

at 25 ^{o}C. This doesn't explain the factor of 50,000. The resolution of this paradox was the finding that IleRS has a separate catalytic site at which it binds valine adenylate that it then hydrolyzes to valine and AMP. This site isn't large enough to bind isoleucine adenylate residues. Thus, if valine adenylate accidentally enters IleRS, it gets tested at the extra binding site first before it is allowed to interact with the bound $tRNA^{Ile}$, at least the vast majority of the time. One might expect that this means that IleRS has two subunits, one for each binding site, but IleRS is a class I aaRS of type α. Many other aaRS's distinguish against noncognate aa's by a similar mechanism. When sufficient selectivity exists, the proofreading mechanism just described isn't necessary. For example, TyrRS (tyrosine aaRS) discriminates between tyrosine and phenylalanine through H bonding to the tyrosine –OH group.





Universality of the genetic code

Variants to the standard code exist in mitochondria that have their own genes for about 13 mitochondrial proteins.

mitochondria	UGA	AUA	CUN	AG $\frac{A}{G}$	CGG
mammals	trp	met		stop	
baker's yeast	trp	met	thr		?
neurospora crassa	trp				?
drosophila	trp	met		ser	
protozoa	trp				
plants					trp
standard	stop	ile	leu	arg	arg

We can rationalize most of these changes: 1) In the standard code, Trp (tryptophan) has a single codon, UGG. UGA is its neighbor also ending with a purine base, and there are 2 other stop signals, UAA and UAG. 2) Met (methionine) has only one codon, AUG. Thus its purine neighbor, AUA, is close and still leaves 2 isoleucine codons. 3) CUN where N is any base for Thr (threonine) in yeast is hard to explain. 4) AGA or AGG being stop codons in mammals still leaves plenty of codons for Arg (arginine). It is close to the Ser (serine) codon as well. 5) CGG for Trp in plants is close to UGG for Trp in the standard code since C is also a pyrimidine. Thus, these variants are for the most part so minor that they can be viewed as evolutionary refinements of a primitive universal code.

The genetic code is highly degenerate, i.e. most amino acids have several codons. This degeneracy is partly attributable to a defect in anti-codon codon recognition called *wobble*. The base in the 5' position of the anti-codon does not always correspond with only one base in the corresponding 3' position of the codon. The base pairing rules for this situation are given by

Wobble

5' anticodon base	3' codon base
C	G
A	U
U	A or G
G	U or C
I	U, C or A

The tRNA's have many modified bases, one of which can be I (inosine) a close, purine relative of A and G. For example, the anticodon for $tRNA^{Ile}$ is UAI while that of $tRNA^{Met}$ is UAC. For isoleucine the code is triply degenerate whereas for methionine the codon is unique. The code's degeneracy has prompted many researchers to imagine that during evolution the code was once a two base code. The two base code that can be extracted from the present code has room for 15 amino acids and one termination signal.

Two base code

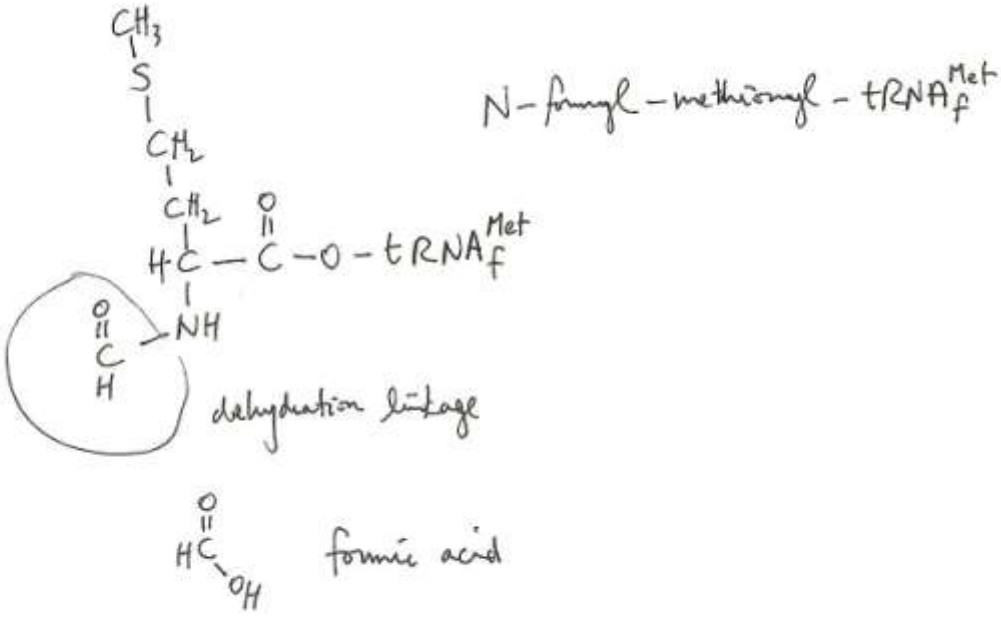
codon	amino acid
5'→3'	
AA	lys
AC	thr
AG	ser or arg

AU	ile
CA	his
CC	pro
CG	arg
CU	leu
GA	asp or glu
GC	ala
GG	gly
GU	val
UA	stop
UC	ser
UG	cys
UU	phe

It has no start codon and is missing 5 amino acids, methionine, tryptophan, tyrosine, asparagine and glutamine. Nevertheless it does contain representatives of every amino acids residue group, especially all of the simple residues. If such a more primitive code is contemplated, it is necessary to assume that the code spacing is still three bases. This would mean that the third base of the codon is completely degenerate. The reason for requiring a three base spacing is so that during evolution there would not also be a required shift from two base spacing to three bases spacing in the mRNA's. It is very difficult to imagine how such a shift could be implemented. As evolution progressed, some third base specificity has evolved and this has allowed a start signal to become incorporated as well as the use of several exotic amino acids.

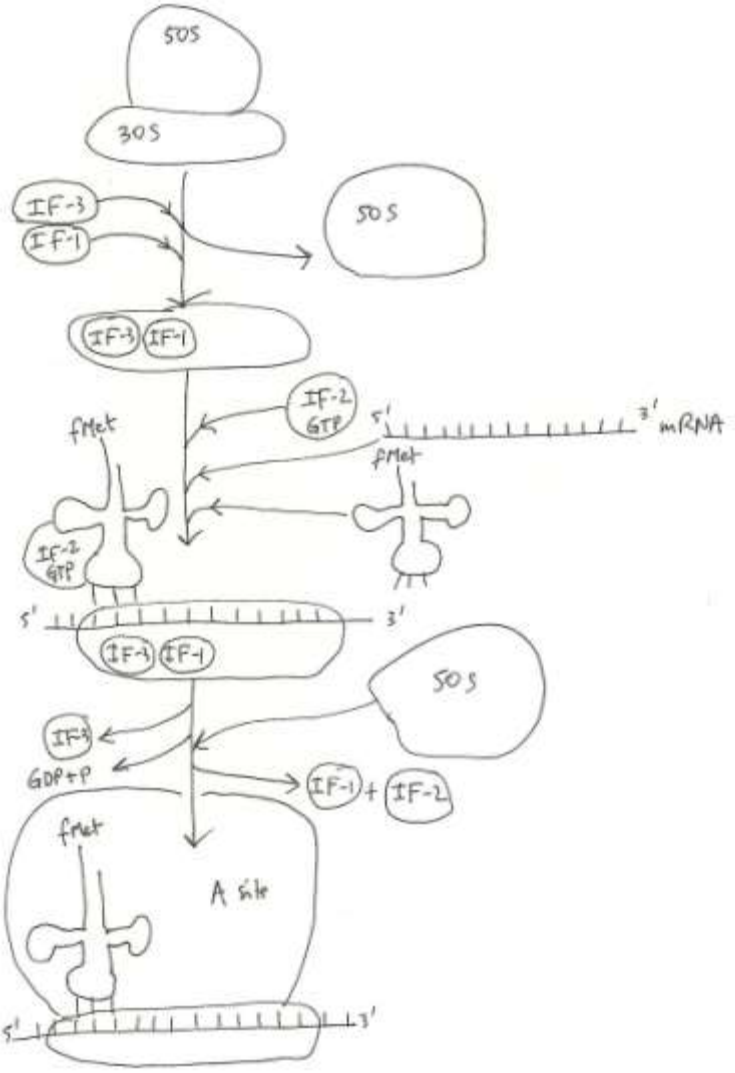
Initiation, elongation and termination of translation

Two different tRNA's read the methionine codon AUG, $tRNA_f^{MET}$ and $tRNA_m^{MET}$. A single MetRS aminoacylates both tRNA's. In the case of $Met - tRNA_f^{Met}$ an enzyme specifically n-formylates the methionine to form $fMet - tRNA_f^{Met}$.

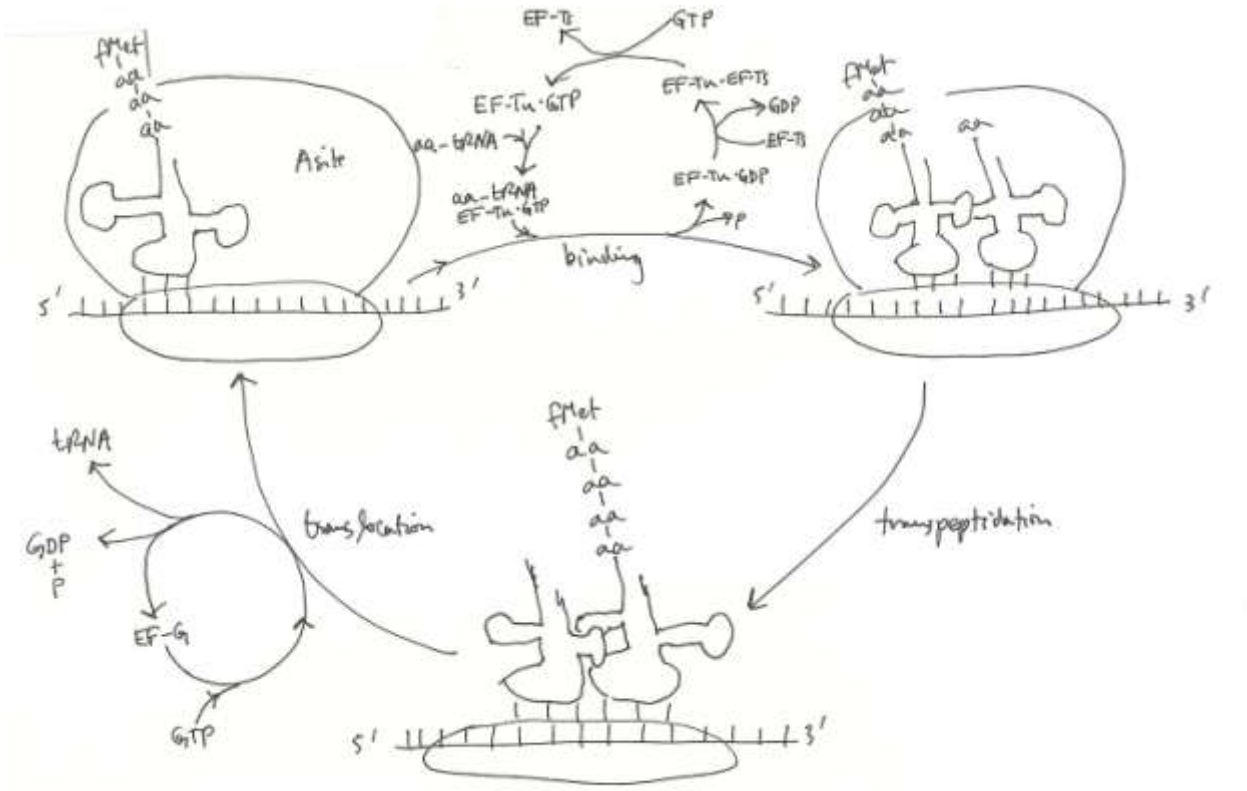


The formylation enzyme does not recognize $Met-tRNA_m^{Met}$. Every protein translation from mRNA is initiated by the codon AUG and the first amino acid is fMet. How does the cellular apparatus know which AUG's are the initiation codons as opposed to simply being codons for interior methionines that are carried by $Met-tRNA_m^{Met}$? In *E. Coli*, the 16S rRNA of the small ribosomal subunit contains a pyrimidine rich sequence at its 3' end that is partially complementary to a purine rich mRNA tract of 3 to 10 nucleotides, the Shine-Dalgarno sequence named after its discoverers, that is located ~ 10 nucleotides in the 5' direction form the AUG start codon. This sequence is found in all known procaryotic mRNA's. Base pairing between the 16S rRNA and the mRNA's Shine-Dalgarno sequence apparently permits the ribosome to select the proper initiation codons from all AUG's.

After a polypeptide has been synthesized, the 30S and 50S ribosomal subunits remain associated as a 70S ribosome. Initiation factor IF-3 binds the 30S subunit and permits dissociation from the 50S subunit. IF-1 accelerates the rate of dissociation. mRNA, and IF-2 in a ternary complex with GTP and $fMet-tRNA_f^{Met}$, subsequently bind to the 30S subunit, in either order. Hence $fMet-tRNA_f^{Met}$ recognition seems not to be mediated by codon anti-codon interaction, the only tRNA-ribosome interaction not requiring such. IF-3 is released and the 50S ribosomal subunit joins the 30S initiation complex. This stimulates IF-2 to hydrolyze GTP to GDP and P. The 30S subunit conformationally rearranges and releases IF-1 and IF-2.

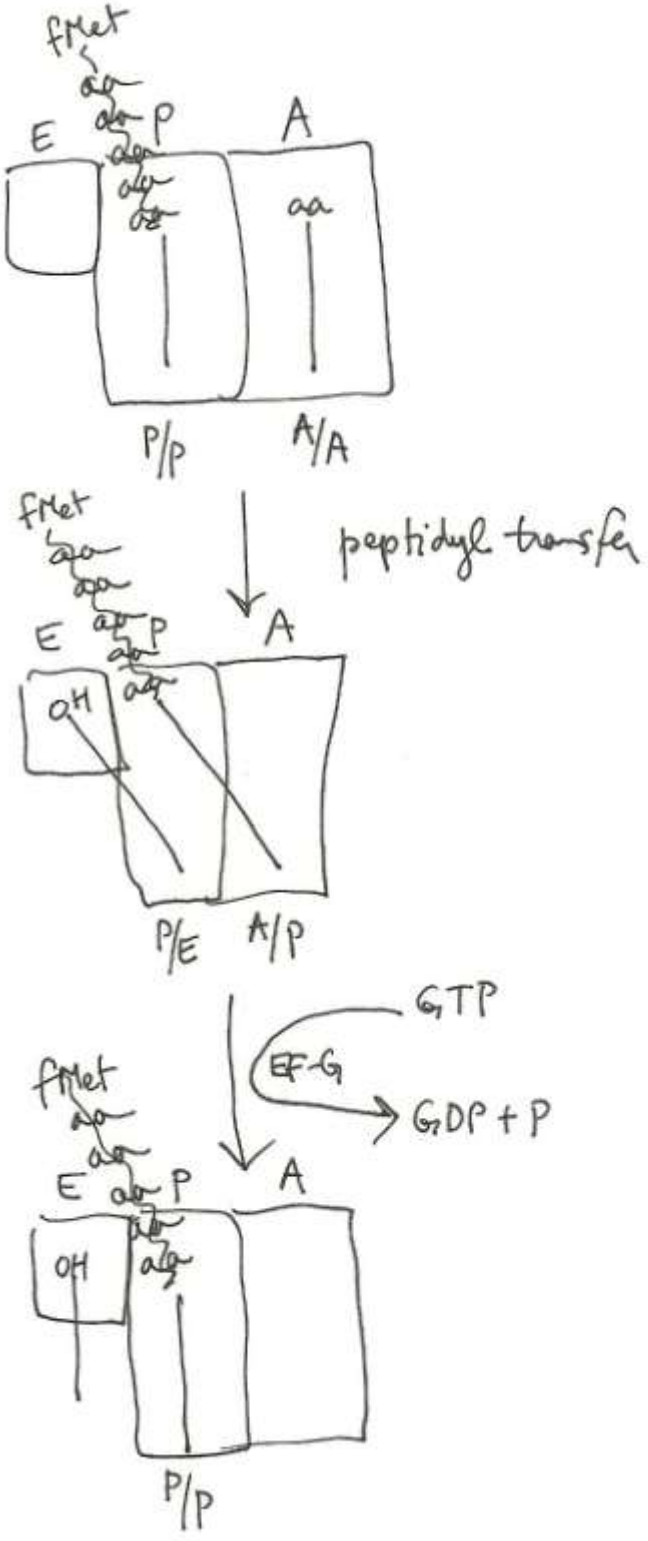


The 50S sununit contains three sites used during elongation of the polypeptide, the E site, the P site and the A site. The $fMet-tRNA_f^{Met}-mRNA-ribosome$ complex has $fMet-tRNA_f^{Met}$ in the ribosome's P site while the empty A site is ready to accept an incoming aminoacyl-tRNA that will be recognized by codon anti-codon base pairing. A binary complex of GTP with elongation factor EF-Tu combines with the aa-tRNA. As this complex binds the ribosome, GTP is hydrolyzed to GDP and P and the aa-tRNA is bound to the A site by codon anti-codon base pairing. EF-Tu-GDP and P are released. EF-Tu-GDP has GDP displaced by EF-Ts, which in turn is displaced by GTP, regenerating EF-Tu-GTP.



Transpeptidation occurs in which the fMet carboxyl forms an amide linkage with the new aa's amino group. EF-Tu is needed to help the new aa-tRNA bind at a rate sufficient to support cell growth. It is present in the cell with ~ 100,000 copies per cell, which is approximately the number of tRNA's. Consequently, almost all the cell's tRNA's are bound to EF-Tu's. But EF-Tu does not bind either $fMet-tRNA_f^{Met}$ or $Met-tRNA_m^{Met}$. The transpeptidation reaction appears to be catalyzed by the 50S subunit's 23S rRNA. This is an example of *ribozyme* activity (Science **289**, 11Aug., 2000). A ribozyme is an RNA that catalyzes a reaction. The repertoire of known ribozyme activities, while impressive, is very much smaller than the repertoire of protein catalysts. Because RNA can serve as a repository of genetic imformation much like DNA does and because it has some ribozyme activities, some researchers have proposed that the origin of life went through a stage called the *RNA world* before there was gene directed protein biosynthesis.

Once the peptide synthesis has commensed, elongation takes place. This process bring into play the E site.



A growing peptide is in the P site and a new aa-tRNA has arrived at the A site. Each is attached to the mRNA by base pairing. A transpeptidation reaction takes the growing polypetide's carboxyl group that is intially attached to the tRNA and joins it to the free amino group of the aa-

tRNA in the A site. The terminal –OH group of the ribose at the 3' end of the tRNA of the P site moves to the E site. The elongated polypeptide attached to the A site tRNA moves to the P site. Elongation factor EF-G binds and hydrolyzes GTP to GDP and P while shifting the mRNA so that the elongated polypeptide's attached tRNA is now entirely in the P site and the freed tRNA is entirely in the E site from which it detaches. The ribosome is now ready for another aa-tRNA to arrive and elongate the polypeptide by one more residue. Notice that the entire growing polypeptide is transpeptidated to the incoming amino acid and not the other way around. To do it the other way around would mean putting the new amino acid on the N-terminus of the growing polypeptide and this terminus is away from the ribosome since it is the C-terminus that is attached to the bound tRNA. It is sterically impossible to have the mechanism work this other way around.

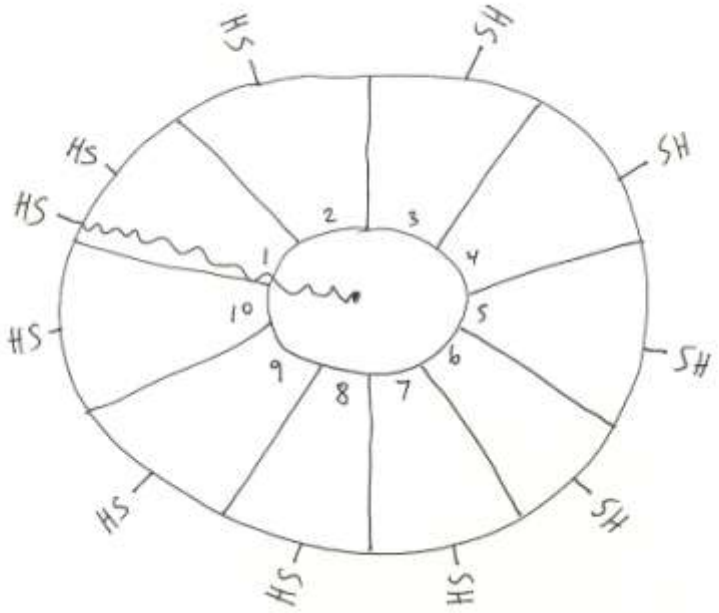
Termination of protein synthesis is signalled by three codons, UAA, UGA and UAG. These codons are recognized by releasing factor. RF-1 recognizes UAA and UAG and RF-2 recognizes UAA and UGA. RF-3 with bound GTP stimulates binding of RF-1 and RF-2. This induces the ribosomal peptidyl transferase to transfer the peptide group to H_2O, thereby releasing it. The uncharged tRNA dissociates from the ribosome. The releasing factors are expelled and GTP hydrolyzes to GDP and P. The mRNA is released producing a ribosome ready for more protein synthesis.

Each of the initiation factors, the elongation factors and the releasing factors is a pure protein. These proteins have evolved to facilitate the steps of protein synthesis. A fundamental problem is to understand the evolution of the entire gene directed protein biosynthesis machinery. How did ribosome, tRNA's, aaRS's and mRNA's evolve? How did the genetic code originate? Why is it a three base code with lots of degeneracy in the third position? Even a hypothetical model may greatly deepen our comprehension of this fantastic process.

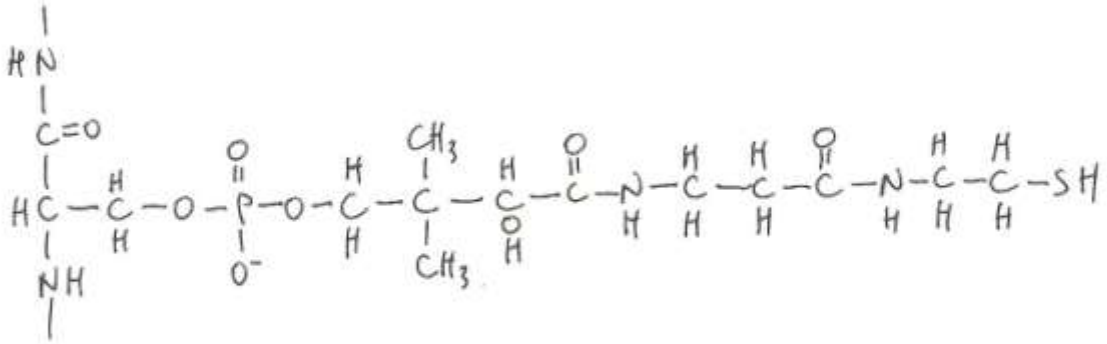
Tyrocidine synthetase

Tyrocidine is a cyclic decapeptide antibiotic synthesized by the bacterium *Bacillus brevis*. It is really a very small protein but it is not synthesized on a ribosome like other proteins. Instead, it is made by a polyenzyme complex.

The complex consists of 4 protein subunits, each of which itself is the product of gene directed ribosomal translation typical of normal protein synthesis. These subunits self-assemble into the structure depicted below.



The central core of the complex is a protein to which is attached a *phosphopantetheine* arm, that terminates in a sulfhydryl group, -SH. This freely rotating arm is 2 nm long. It is probably linked to a core protein serine residue

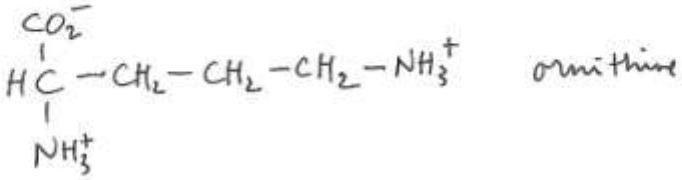


since this is how it is linked in fatty acid synthetase, another large enzyme complex. The core protein has MW ~ 20,000 d, and is tightly bound to a polyenzyme having 6 sites numbered 5, 6, 7, 8, 9 and 10 that has MW ~ 440,000 d. A smaller polyenzyme containing sites 2, 3 and 4 is complexed to the larger component. It has MW ~ 230,000 d. Finally, a smallest enzyme subunit containing only site 1, with MW ~ 100,000 d, completes the structure. Each numbered site contains a -SH group, probably a cysteine residue. Each site is specific for a particular amino acid. This specificity is given in the table.

Tyrocidine site sequence

site	amino acid
1	D-phe
2	pro
3	phe
4	D-phe
5	asn
6	gln
7	phe
8	val
9	orn
10	leu

Two features of this table are peculiar. *D-phenylalanine* appears twice instead of the usual L-phenylalanine. In fact, L-phenylalanine is attached and then *racemized* to the D form by the enzyme site. Ornithine is used although it is not found at all in gene directed proteins.

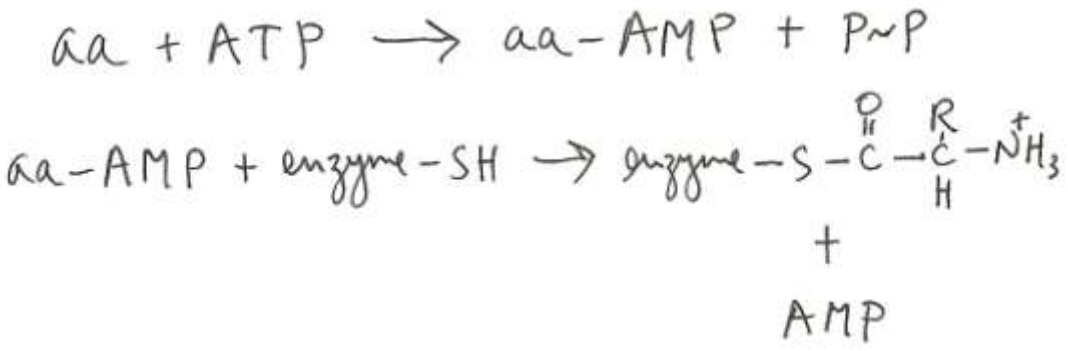


lysine has $-(CH_2)_4-$

arginine + H_2O → ornithine + urea

This amino acid is similar to lysine which has a chain of 4 -CH_2-‘s instead of just 3. It is also the hydrolysis product of the amino acid *arginine*, along with *urea*. Site 9 simply recognizes ornithine instead of lysine or arginine, probably by virtue of a recognition site that is too small for these closely related but larger amino acids.

Attachment of the amino acids requires prior activation with ATP to form the activated aminoacyl adenylates that then react with the -SH groups to form *thiocarboxyl esters*.



The adenylates are in solution, unlike their situation when they are intermediates for gene directed protein synthesis in which the remain enzyme bound. The thioesters maintain enough free energy of activation to promote peptide formation. Indeed, it was once thought that these thioesters might represent an early evolutionary pathway to protein synthesis before the genetic apparatus evolved. This may be so, but the entire complex of proteins here is the product of genes.

Tyrocidine synthesis begins after all 10 sites are esterified with their specific amino acids. The D-phe on site 1 transfers to the proline at site 2 to form a thioesterified dipeptide, N-terminal D-phe-pro-S-enzyme. When this dipeptide transfers to the phe at site 3, the rest of the polymerization is triggered until the decapeptide, N-terminal D-phe-pro-phe-D-phe-asn-glu-phe-val-orn-leu-S-enzyme is formed. At this point, cyclization occurs as a peptide bond forms between the N-terminal amino group of D-phe and the carboxyl group of leu that is attached to the enzyme complex by a carboxyl thioester at site 10.

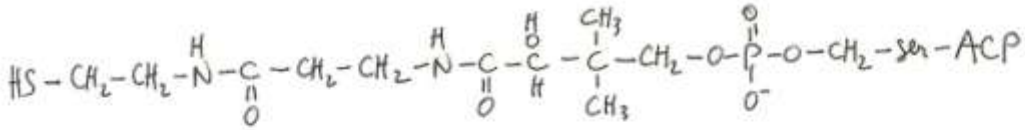
The facsinating mechanism of elongation from site 3 through site 10 involves the phosphopantetheine arm. The peptide at any one of these sites gets to the next site by first being *transthiolated* to the -SH group of the rotating phosphopantetheine arm. The arms swings to the next site and the peptide *transpeptidates* onto the amino group of the amino acid at that site. Then this elongated peptide transthiolates back onto the phosphopantetheine arm and is carried

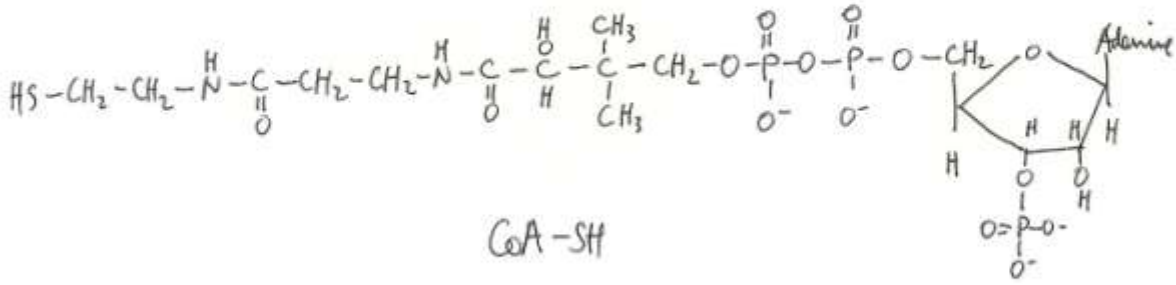
to the next site. The motion of the arm is caused by thermal agitation known as Brownian motion.

Fatty acid synthesis

The synthesis of fatty acids, such as palmitic acid, also utilizes a rotating phosphopantetheine arm. Seven different enzymatic reactions are involved. In *E. Coli*, these reactions are catalyzed by independent enzymes, in chloroplasts it is similar (more evidence that chloroplasts are originally cyanobacterial symbiots), in yeast fatty acid synthetase is a 2,500 kd $\alpha_6\beta_6$ multifunctional enzyme and in animals it is a 534 kd multifunctional enzyme consisting of 2 identical polypeptide chains.

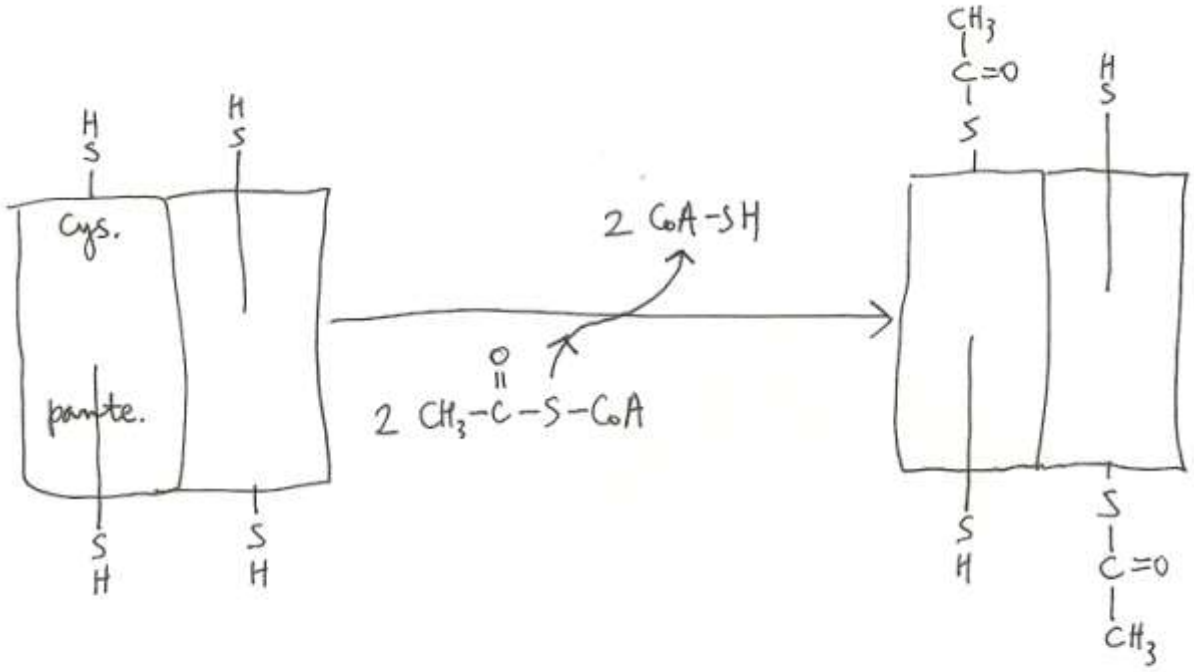
The reactions needed for fatty acid synthesis are sequentially occuring on a rotating phosphopantetheine arm. In the fatty acid synthetase complex, the phosphopantetheine is covalently linked to a serine residue in the *acyl carrier protein* ACP. In CoA, the same arm is attached to AMP instead.



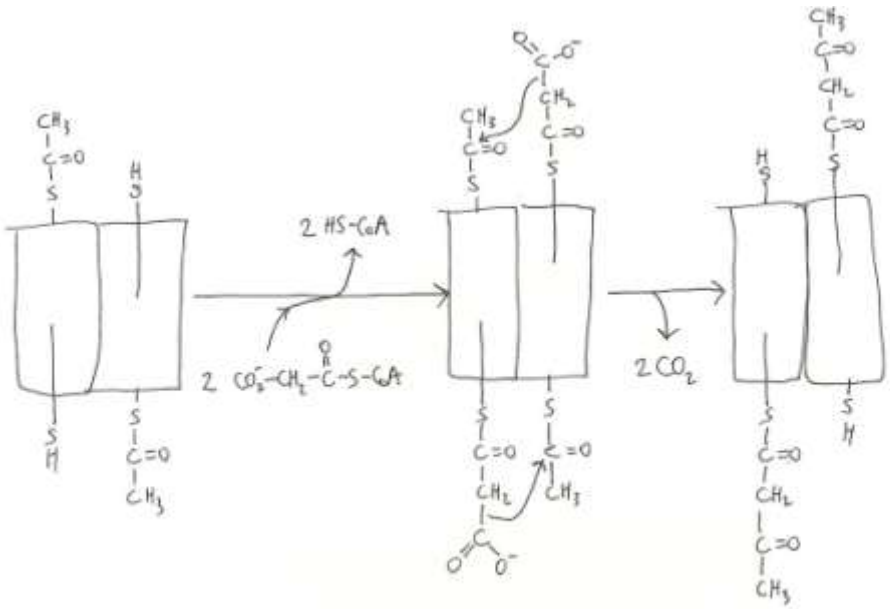


The enzyme also contains a cysteine residue that is important in the mechanism. The phosphopantetheine arm carries the substrate from one reactive site to another.

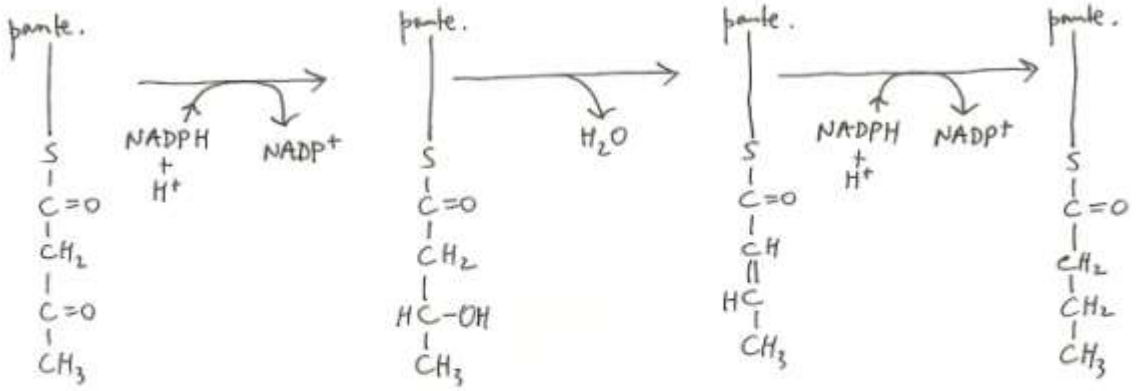
The first step is a *transthioesterification* of acetyl from acetyl-CoA to the -SH of cysteine in the enzyme complex. Since the complex is a dimer in animals, this occurs at two symmetrical sites.



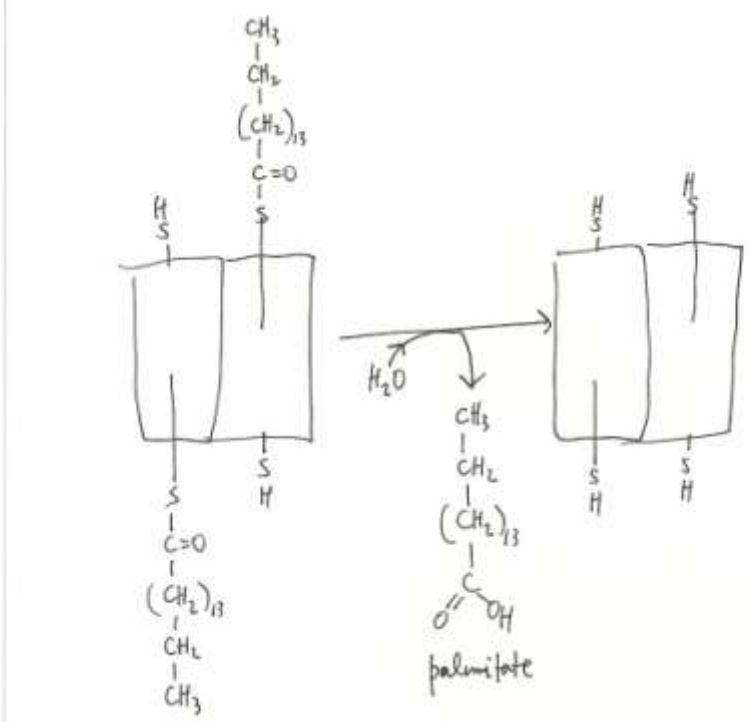
The second step is the transthioesterification of *malonyl* from malonyl-CoA to the -SH of phosphopantetheine.



There then follows a reduction by NADPH, a dehydration, another NADPH reduction and a transesterification of the growing fatty acid from phosphopantetheine to cysteine.

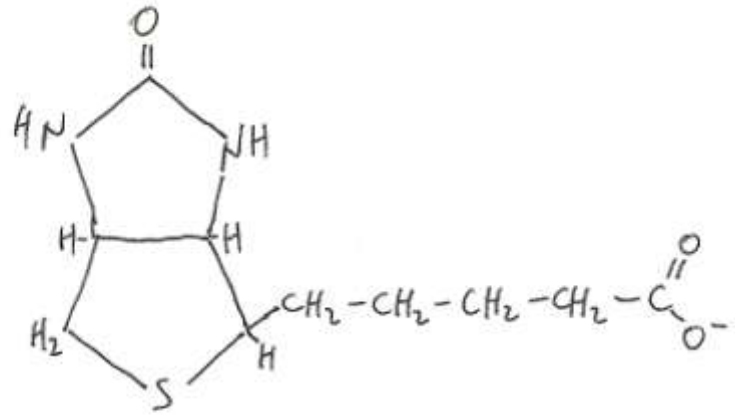


More malonyl-CoA comes in to form another malonyl-phosphopantetheine thioester. This again leads to a decarboxylation and condensation of the growing fatty acid onto the malonyl-phosphopantetheine (minus the CO_2). This sequence continues for 6 cycles until palmityl-phosphopantetheine is formed and palmitate is hydrolytically released from the enzyme.



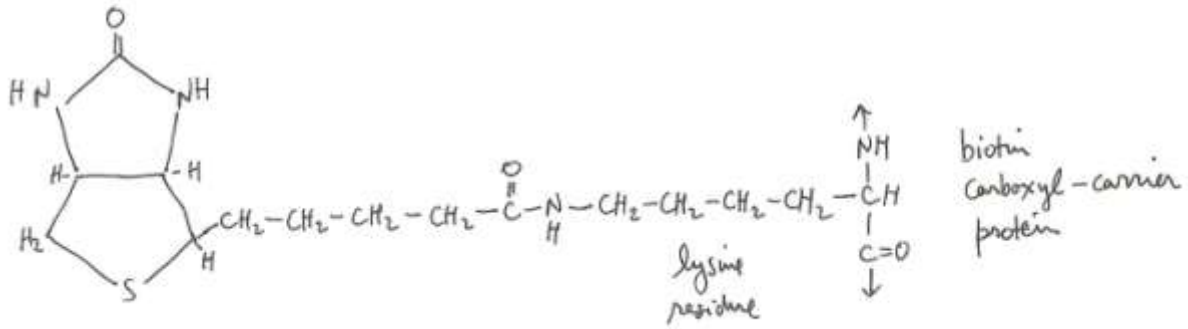
Note that as in protein synthesis, the growing polymer is transferred to the incoming monomer, not the other way around.

Where does the malonyl-CoA come from? The coenzyme, *biotin*, is the active center of the enzyme for making malonyl-CoA.



This prosthetic group also occurs in pyruvate carboxylase, where oxaloacetate is made, and in propionyl-CoA carboxylase that occurs in fatty acid degradation. Biotin forms an amide linkage

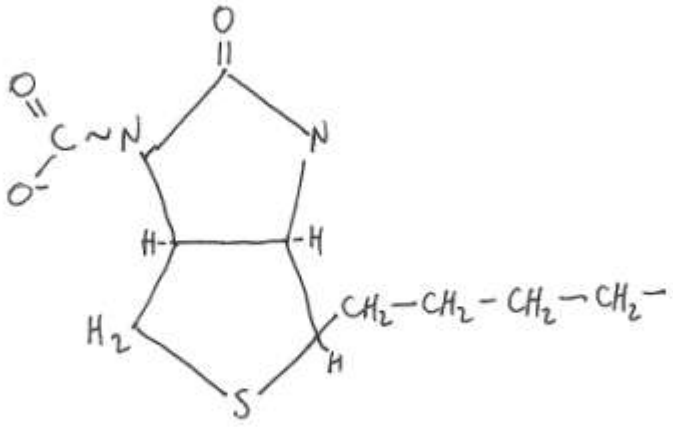
with the ε-amino group of a lysine residue in the biotin carboxyl-carrier protein, part of the enzyme complex. This combination of biotin and lysine is called *biocytin*.



The reaction steps are the creation of an activated carboxyl group and the carboxylation of acetyl-CoA with it to make malonyl-CoA.

$$biotin-enzyme+ATP+HCO_3^- \rightarrow CO_2 \sim biotin-enzyme+ADP+P$$
$$CO_2 \sim biotin-enzyme+acetyl-CoA \rightarrow malonyl-CoA+biotin-enzyme$$

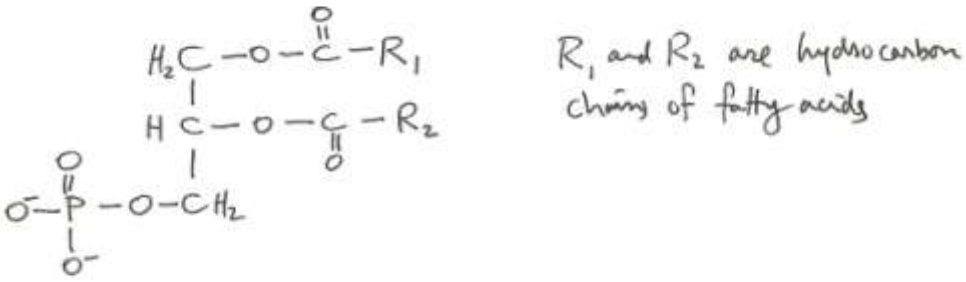
The energy of ATP is required for activation. Carbonate, HCO_3^-, the aqueous form of CO_2 is the source of the carboxyl group. The chemical form of the active carboxyl carrier intermediate is a modified biotin.



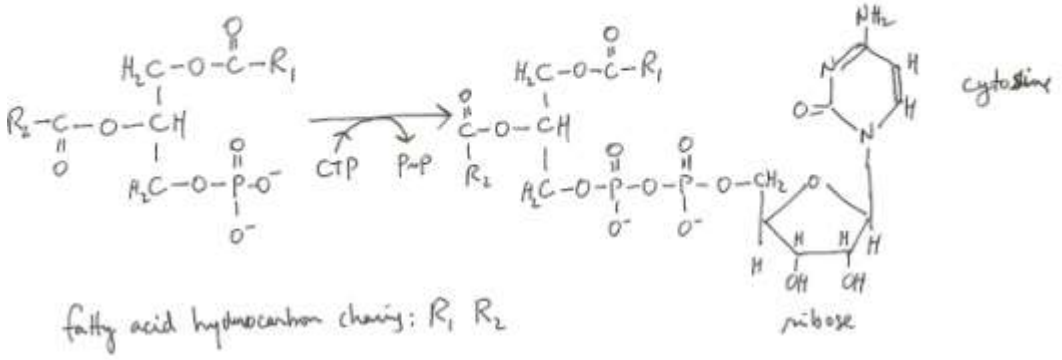
The first step is catalyzed by *biotin carboxylase* and the second by *transcarboxylase*. These two enzymes are in a complex with the biotin carboxyl-carrier protein and the biocytin arm moves back and forth between the two reaction sites by thermal Brownian motion.

Glycerophospholipid synthesis

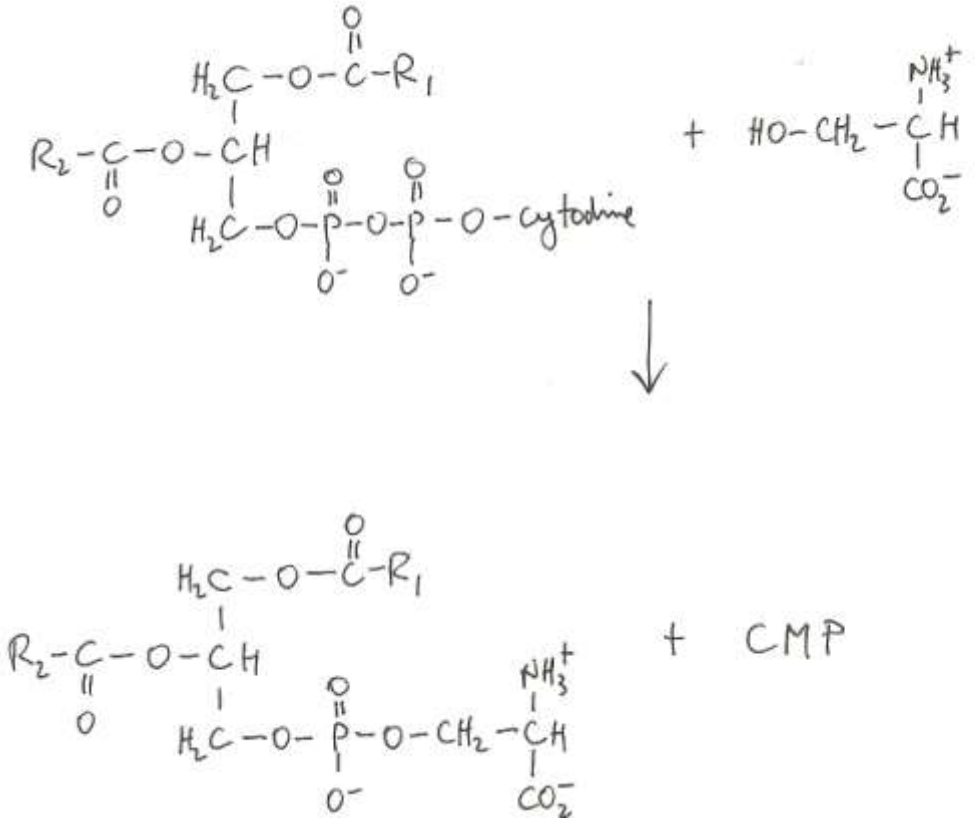
In figure below diacylglycerol-3-phosphate is dipicted. The majority of lipids in membranes are derivatives of this compound in which the hydroxyl group of an alcohol combines with the phosphate to form an ester linkage.



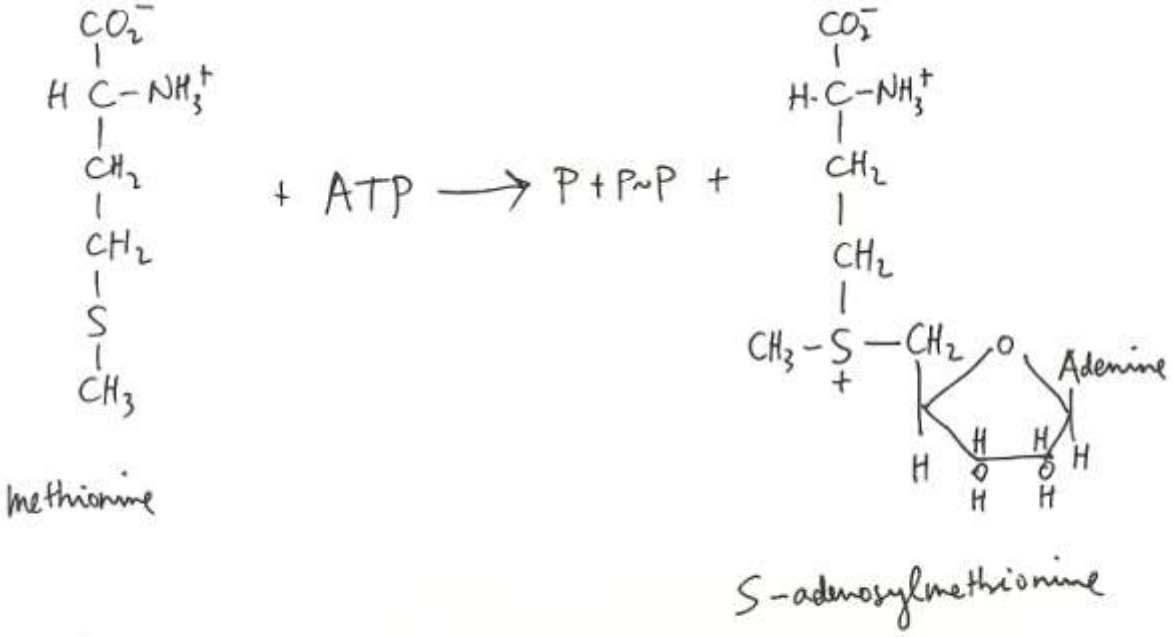
To accomplish this dehydration condensation, acitvation energy is provided by CTP (cytodine triphosphate) that contains the base cytosine. CTP combines with diacylglycerol-3-phosphate to form CDP-diacylglycerol and pyrophosphate.



This activated intermediate can react with serine to form *phosphatidyl serine* and CMP (cytodine monophosphate).



A series of subsequent modifications lead to *phosphatidyl ethanolamine* and *phosphatidyl choline*. The methylation step requires an activated form of methyl groups that are supplied by *S-adenosylmethionone*.



This compound is synthesized from methionine and ATP. This sctivation reaction is unusual in that all three phosphate groups in ATP are released at the end of the activation. Once again, the key role of a sulfur atom is seen.

While the polymerization of sugars is left to other sources, suffice it here to observe that activation in this case requires UTP (uridine triphosphate) based on the base uracil. Thus ATP is used for amino acid activation, CTP is used for activation in lipid synthesis and UTP is used for saccharide activation. Of course ATP, CTP, UTP and GTP are all required for RNA synthesis but CTP, UTP anf GTP become triphosphates by virtue of the supply of ATP generated by metabolism.

Information density

The advent of modern molecular genetics has lead many scientists to the notion that life is driven by information. The crucial importance of energy in its many forms has already been emphasized in this text. At this juncture, the relative importance of information will be considered.

Recent technological advances have led to predictions of optical tapes with information densities of 2 bits per micron and of HDD magnetic tapes with 14 bits per micron (350,000 bits per inch). How does this compare with DNA and proteins?

If the probability of an event is P, then the information, I, associated with that event occurring is defined to be

$$I = -\log_2 P$$

When $\log_2$ is used, it is said that I is in *bits*. Thus, the information associated with rolling a 5 on a fair die is

$$I = \log_2 6 = 2.58\, bits$$

A DNA molecule is made up of N base pairs, of which there are 4 types. This means the information content is 2 bits per base pair ($\log_2 4$). The total number of possible DNA's of N base pairs is 4^N. The information content to make one of them is 2N (i.e. $\log_2 4^N$).

The entropy change in an event happening with probability P is given by

$$\Delta S = k_B \ln P = k_B \ln 2 \log_2 P = -0.693\, k_B I$$

The spacing of base pair in DNA is about 3 angstroms per base pair. This implies an information density of 2 bits per 3 angstroms or 6.7 x 10^3 bits per micron. At 25 ^{o}C, the contribution from information alone to the free energy cost of making a DNA with N base pairs is

$$-T\Delta S = 0.693\, k_B TI = N \times 5.54 \times 10^{-14} ergs$$

A phosphodiester bond costs about 12 kcal per mole which translates into 1.66 x 10^{-12} ergs per base pair. This is roughly 30 times the information cost!

A similar calculation can be done for proteins. Each amino acid takes up about 5 angstroms. This implies 4.32 bits ($\log_2 20$) per 5 angstroms or
8.6 x 10^3 bits per micron. The information cost at 25 ^{o}C per amino acid is

$$-T\Delta S = 0.693\, k_B TI = 12 \times 10^{-14} ergs$$

A peptide bond costs about 12 kcal per mole for the activation of the amino acid to an aminoacyl adenylate. This is 8.3 x 10^{-13} ergs per bond. This is only a factor of 7 times the information cost. However, every peptide bond made on a ribosome also requires 2 GTP's for tRNA binding and translocation on the ribosome. This brings the total cost of synthesis to 21 times the information cost. The synthesis of the molecules themselves is clearly expensive, both for DNA and proteins, and information considerations are quite minor. Evolution has proceeded because of energy and information has accumulated as a characteristic of fitness.