Searching for Molecular Solutions – **Cited Notes**

CHAPTER 5

These Files contain details on all references to this ftp site within **Chapter 5** of *Searching for Molecular Solutions*. The page numbers of the book where the reference is made are shown in the Table below, the corresponding page number for this file, and the title of each relevant section.
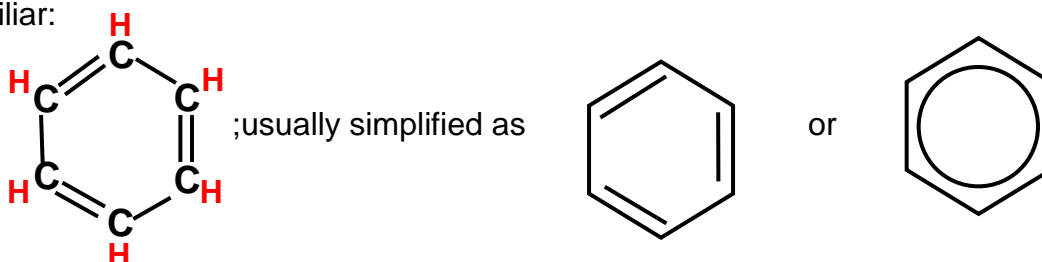
Contents:

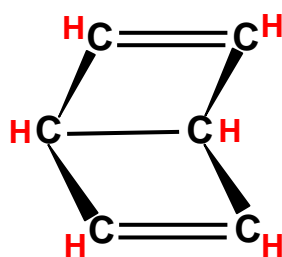| Book Reference Page Number	Page Number in this File	Section	
		No.	**Title**
158	2	12	Stereochemistry
178	5	13	Specialized Ribosomes
184	6	14	Genetic Code Expansion
185	13	15	Expanded Genetic Code Applications
186	18	16	Specialized Ribosomes

Section 12: ***Stereochemistry***

Cited on p. 158 of *Searching for Molecular Solutions*

Basic chemistry, especially the chemistry of carbon compounds, includes the concept of isomers, as a term for compounds with the same molecular formula but with non-homologous alignments of component atoms. As an example, the simple formula C_6H_6 immediately brings the aromatic hydrocarbon benzene to mind for anyone with a basic background in organic chemistry. In turn, the ring structure of benzene discovered by Kekulé in the mid-nineteenth century is very familiar:
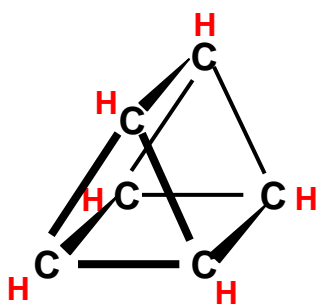
;usually simplified as ... or

However, 217 non-benzene C_6H_6 isomers have been proposed using normal carbon-hydrogen valencies, although most are highly strained and less than 40 have actually been derived [1].

Two examples are 'Dewar benzene':

and prismane:

Both satisfy the C_6H_6 formula and the rules for carbon valency, but both are much less stable than benzene itself.

Yet the tally of benzene isomers rises even higher (to 328[1]) if one includes the special isomeric category of *stereoisomers*.
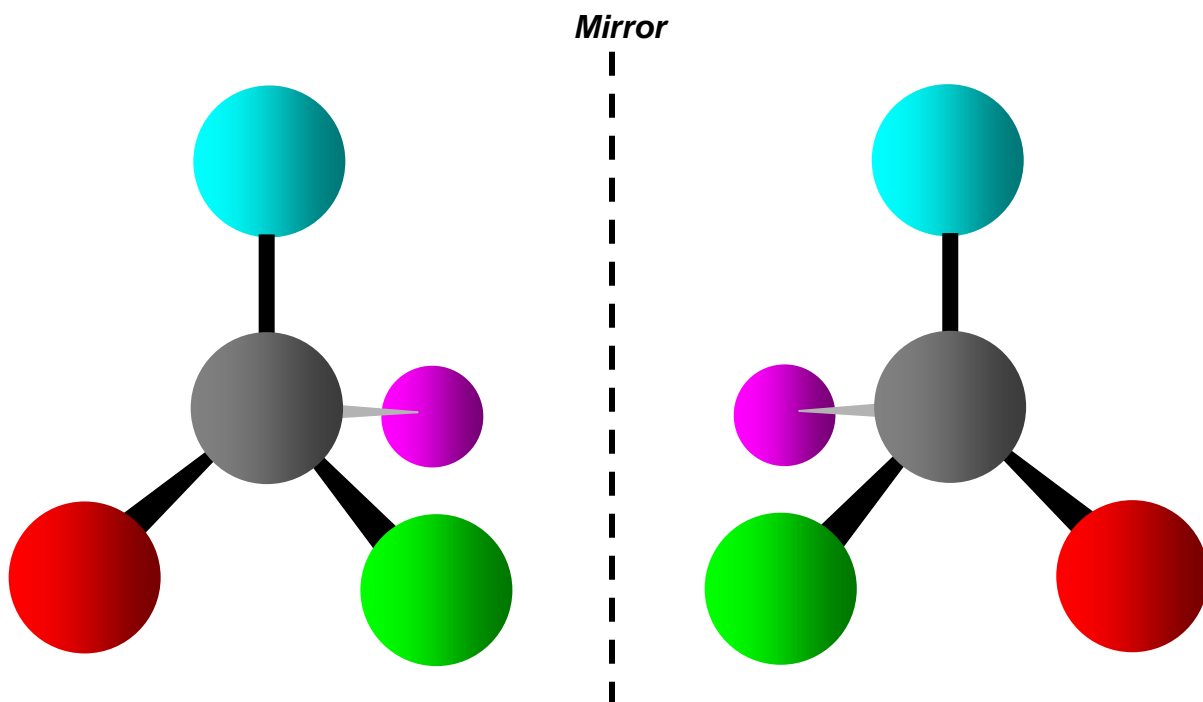
Mirror

Fig. 5.Na

Representation of an asymmetric carbon atom (gray ball) with four different substituents in a tetrahedral shape, showing that its mirror image is non-superimposable.

Given that carbon has a valency of four, with preferred bond angles in a tetrahedral shape, it logically follows that with four different substituents, isomers can exist which are mirror images of each other, as shown above (Fig. 5.Na). These stereoisomers are termed *enantiomers*, and have the property of rotating plane-polarized light, or optical activity. A compound is defined as dextrorotary (as with D-sugars, such as the D-ribose found as a component of nucleic acids) through conferring a rightward rotation of a plane-polarized beam and levorotary

(as with L-amino acids) if the rotation is in the leftward direction. (Other stereochemical nomenclature systems also exist [2]). Carbon atoms in a compound with four different substituents are termed asymmetric or *chiral* centers. Although pairs of enantiomers have the same physical properties, stereochemistry in general is critically important in biological molecular recognition, a recurring theme in *Searching for Molecular Solutions*. There is vastly more that could be said on this and related topics, far beyond the scope of this simple introduction; a basic organic chemistry text (for example; the classic Morrison and Boyd [3]) is good starting point for further exploration for those requiring more background.

Section 13:   *Specialized Ribosomes*

Cited on p. 178 of *Searching for Molecular Solutions*

This reference to specialized ribosomes is covered within a later ftp citation for this chapter.

Thus:

Section 14:   ***Genetic Code Expansion***

Cited on p. 184 of *Searching for Molecular Solutions*

Positive-negative selections for genetic code expansion

This section provides additional details on selection procedures used (largely by the laboratory of Peter Schultz) in the pathway towards expanding the natural genetic code.

At the beginning, what is the best codon for reassignment? Initial work in this area used *E. coli* as the host, and amber UAG codons were chosen since these are the least frequent stop codons in this organism (only 7.6% of all E. coli translational stops, from annotated *E. coli* sequence [4]). Pilot experiments found that reassignment of endogenous *E. coli* tRNAs and aminoacyl-tRNA synthetase specificities towards 'orthogonal' status was difficult, so an alternative starting pair was used from an archaean organism. (Archaean tRNAs are usually poorly charged by *E. coli* aminoacyl-tRNA synthetases, but show improved translational function in *E. coli* compared to eukaryotic tRNAs [5]). An archaean tRNA for a specific amino acid can then be altered into an amber-suppressor by changing its anticodon. (This has been done for an archaean tyrosine-tRNA, analogously to the suppression depicted in Fig. 5.7 of *Searching for Molecular Solutions*).

But this anticodon change might also alter the recognition of the mutant suppressor tRNA by its cognate synthetase [6], although for archaean tRNAs the anticodon region contribution to such recognition is believed to be small [7]. In any case, directed evolution can be applied towards deriving orthogonal suppressor tRNAs and aminoacyl-tRNA synthetases by means of appropriate selections,

both negative and positive. The first step is to ensure that the chosen tRNA is truly orthogonal in the desired host, such that it is both translationally functional and charged specifically by its cognate aminoacyl-tRNA synthetase. Since it is necessary to preserve tRNA recognition by its cognate synthetase, random mutagenesis should logically be focused on regions in the target tRNA molecule *excluding* those known to directly interact with its cognate synthetase enzyme [5]. From such a mutant library, specific tRNAs which are translationally functional as amber suppressors and charged only by the cognate aminoacyl-tRNA synthetase can be 'pulled out' by application of the right selective pressures. The most powerful way of achieving this is the combination of both positive and negative selections, sometimes referred to as a 'double sieve' strategy [8]. A negative selection is designed to delete cells lacking a particular function, while cells possessing a desired function can be obtained with the appropriate positive selection. In the specific instance we are concerned with here, negative selection can be applied by expressing the tRNA library in the presence of:

(Lethal gene with amber mutations) / *without* cognate aminoacyl-tRNA synthetase:
If cells grow, then → $tRNA_{CUA}$ (amber anticodon tRNAs / *nonchargeable* by host synthetases)

If any UAG amber suppressor tRNA is functional, then expression of the lethal gene will proceed and the host cell will die. But this can only occur if any of the host aminoacyl-tRNA synthetases charges members of the tRNA library, so only library members 'invisible' to the host enzymes allow cell survival. The resulting selected tRNA library fraction can then be subjected to a positive selection process by expressing them in the presence of:

(Antibiotic resistance gene with amber mutations) / *with* cognate aminoacyl-tRNA synthetase; in the presence of the corresponding antibiotic:

> If cells grow, then → tRNA$_{CUA}$ (amber UAG codon-suppressing and orthogonal; that is to say, it must be charged *only* by cognate synthetase)

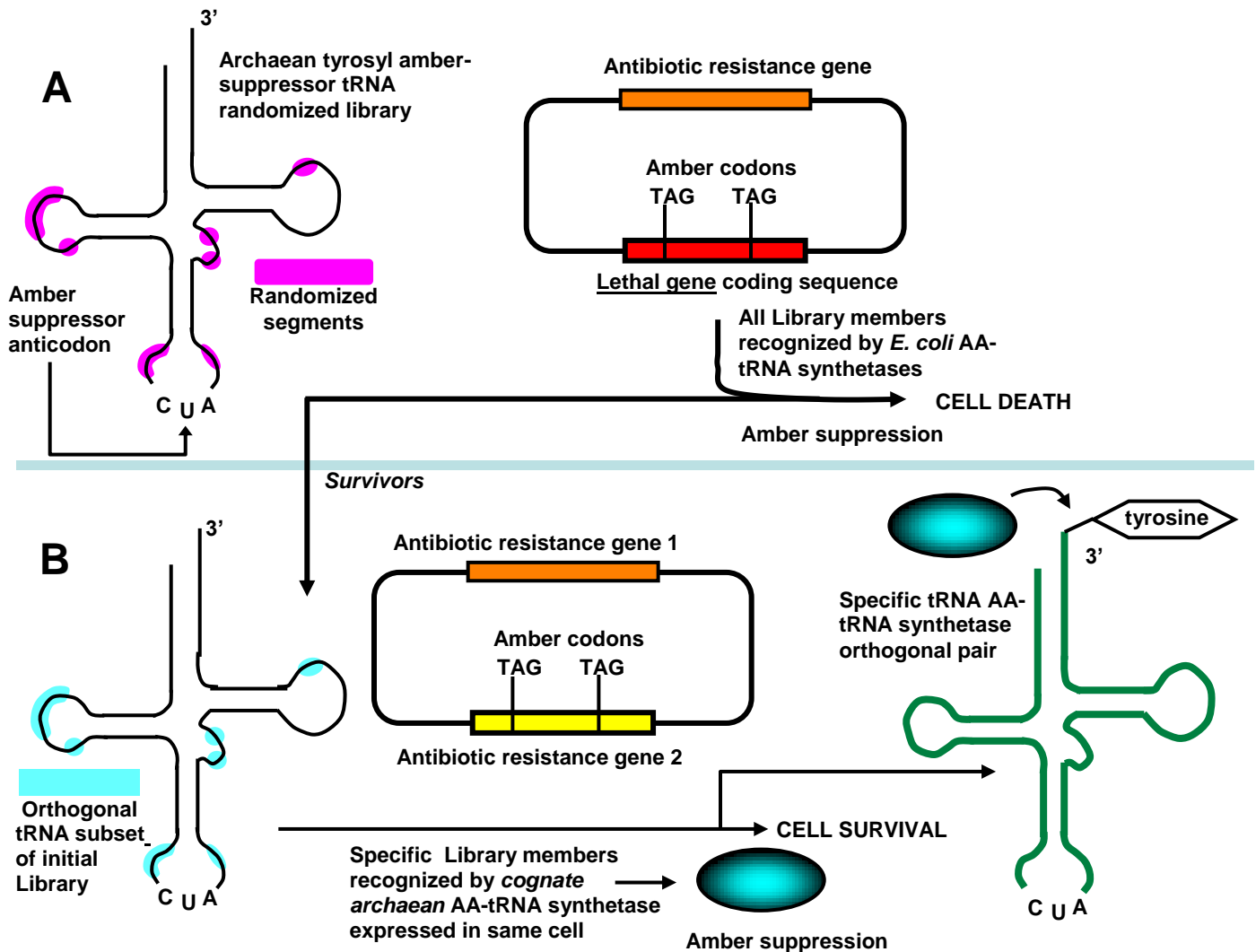These negative and positive selections are schematically depicted in Fig. 5.Nb.

Fig. 5.Nb

Selections to achieve orthogonality of tRNA / and cognate aminoacyl -tRNA synthetases [5]. **A.** Negative selection for tRNAs not recognized by any *E.coli* aminoacyl (AA)-tRNA

synthetases. The anticodon from an archaean tyrosyl tRNA is changed to an amber suppressor triplet, and other specific regions not important for cognate aminoacyl-tRNA synthetase recognition are randomly mutated. The resulting directed library is then expressed in an appropriate *E. coli* host lacking any endogenous suppressor tRNAs. Any functional members of this tRNA library will enable suppression of amber stop codons (TAG at the DNA level) within a gene encoding a lethal product, when a plasmid bearing this sequence configuration is co-transformed into the same cells. Therefore, in this situation cells will be killed if they bear a specific library member which can be charged by endogenous *E. coli* aminoacyl-tRNA synthetases, and thus a negative selection is exerted in favor of orthogonal tRNAs. **B.** Positive selection for tRNAs from within the surviving library subset of (A). Here the library subset is co-expressed in the presence of the cognate (archaean) aminoacyl-tRNA synthetases and also a plasmid where amber mutations are present in a positively-selectable gene encoding antibiotic resistance. Final specific tRNAs are therefore only recognized by their cognate aminoacyl-tRNA synthetases and not by corresponding *E. coli* enzymes.

In this case, for cell survival, suppression of the amber stop codons within the antibiotic resistance gene is required. If the only UAG codon-recognizing suppressor tRNA available is from within the above sub-library, then only tRNAs which retain the ability to be charged by the original cognate synthetase (with its original natural amino acid) will allow survival and growth. (Remember that only tRNAs which *cannot* be charged by the normal host synthetases are represented in the sub-library). The overall result is an 'orthogonal' tRNA$_{CUA}$ (amber-stop codon-suppressing) / synthetase pair which are entirely specific for each other in the foreign host cell (Fig. 5.Nb).

All to the good, but this does not yet bring in the desired unnatural amino acids. For this, aminoacyl-tRNA synthetases themselves must be tinkered with. Although these enzymes can be grouped into two major classes based on sequence and structural differences, they all have comparable catalytic mechanisms [6]. A feature which facilitates their engineering and directed evolution

is their possession of discrete binding regions for target tRNAs, amino acids, and ATP (a required cofactor) [6,9], and this structural knowledge allows mutagenesis and rounds of selection to be targeted at sites appropriate to the changes sought. Consequently, for altering any specific aminoacyl-tRNA synthetase towards utilizing an unnatural amino acid, the normal amino acid-binding site is a logical first choice. The combination of this rational decision with directed evolutionary approaches for the selection of desired variants has paid off. Once again, it is important to implement powerful positive and negative selections. Applying positive selection to find functional enzymes, the aminoacyl-tRNA synthetase mutant library can be expressed in the presence of:

> (Antibiotic resistance gene with amber mutations) / tRNA$_{CUA}$ (amber-suppressor; orthogonal) / *With* desired unnatural amino acid; in the presence of the corresponding antibiotic:
>> If cells grow, then → sub-library of synthetases which can use the desired unnatural amino acid or any natural amino acid ♥.

Remember that previous selections have ensured that only the cognate foreign synthetase (and no host synthetase) can act upon the evolved suppressor tRNA (tRNA$_{CUA}$). Yet only cells where the exotic suppressor tRNA is charged can make the antibiotic and survive. In this situation the cognate synthetase for the suppressor tRNA is in the form of a site-directed, but randomized, amino acid-binding library of variants within which *any* natural *or* unnatural binding specificity might theoretically be found. In the cellular environment in the above selection arrangement, all natural amino acids and only the provided specific unnatural amino acid are present. Therefore, by this combination of factors we are selecting for variants within the synthetase library which can charge the suppressor tRNA with any natural amino acid or the provided unnatural amino acid.

♥ This is true if the antibiotic resistance gene is tolerant ('permissive') for a wide range of amino acid substitutions at the sites of the amber stop mutations.

A negative selection can then remove synthetases specific for natural amino acids. The functional synthetase sub-library is then expressed in the presence of:

> (Lethal gene with amber mutations) / tRNA$_{CUA}$ (amber-suppressor; orthogonal) / *Without* desired unnatural amino acid
> > If cells grow, then → sub-library of synthetases which cannot use natural amino acids [♥].

Any aminoacyl-tRNA synthetases which can charge the suppressor tRNA will enable translation of the lethal gene, and the host cell will die. So the remainder are in theory only synthetases which can utilize the provided unnatural amino acid. In practice, repeated rounds of this positive / negative selection process is necessary to obtain the desired synthetase specificity, but it has been successful for a large number of different unnatural amino acids [5,10]. The resulting aminoacyl-tRNA synthetase capable of charging its cognate tRNA$_{CUA}$ with an unnatural amino acid represents a true extension of the genetic code *in vivo*.

But the ability of a host cell to make use of such an extended code is limited by the availability of the unnatural amino acid itself, which can be provided in the external medium as long as it is transportable into the cell (as in Fig. 5.10 of *Searching for Molecular Solutions*). In principle, the host cell can also be equipped with the enzymatic machinery to synthesize the novel amino acid itself, as has been shown with the natural genes for pyrrolysine and its genetic encoding in *E. coli* [11]. In an analogous but 'unnatural' manner , *E. coli* host cells have been engineered to both synthesize the nonstandard amino acid p-aminophenylalanine and encode it genetically, such that it is inserted at amber codons [12]. By this feat, the host cells achieved autonomous usage of the novel

---

[♥] Analogously with the antibiotic positive selection, this applies if the lethal gene can accept any natural amino acid at the site of the amber stop codons and retain the lethal phenotype.

amino acid, and this work enhances the future prospects for sophisticated evolutionary studies on the advantages of an expanded genetic repertoire [12].

Section 15: *Expanded Genetic Code Applications*

We can group applications for genetic code expansion into a limited number of broad categories (Fig. 5.Nc). For some purposes, proteins with unnatural amino acids at predetermined sites (by virtue of codon reassignment) have immediate utility. Most of these can be described as 'labels' of one type or another. Protein fluorescence has been extremely valuable for a huge range of studies, but the vast majority of proteins are not naturally fluorescent and require the addition of a fluorescent label. A popular means of doing this has been the genetic tagging of proteins of interest with Green Fluorescent Protein (GFP) or other fluorescent polypeptides [13,14]. Although these natural or engineered fluorescent tools are not large by protein standards, they still can generate unwanted perturbations for some *in vivo* measurements. The alternative of direct labeling with small chemical fluorophores is limited to *in vitro* use or involves the introduction of acceptor sites into natural proteins [15]. GFP itself has been engineered for altered fluorescent properties by the introduction of unnatural amino acids [16], but insertion of an unnatural fluorescent amino acid at any predetermined site for any protein has clear theoretical advantages. The appropriate code reassignments and synthetase evolution for such purposes has been achieved for yeast or *E. coli* [15,17]. In mammalian cells, 5-hydroxytryptophan has been applied as a genetically encoded unnatural amino acid with fluorescent and other useful labeling properties [18]. In this study, a prokaryotic tryptophanyl-tRNA / aminoacyl-tRNA synthetase pair was used as the source material for deriving an appropriate orthogonal system, and only a single residue change in the amino acid substrate site of the synthetase was sufficient for altering its specificity from tryptophan to the 5-hydroxy derivative [18].
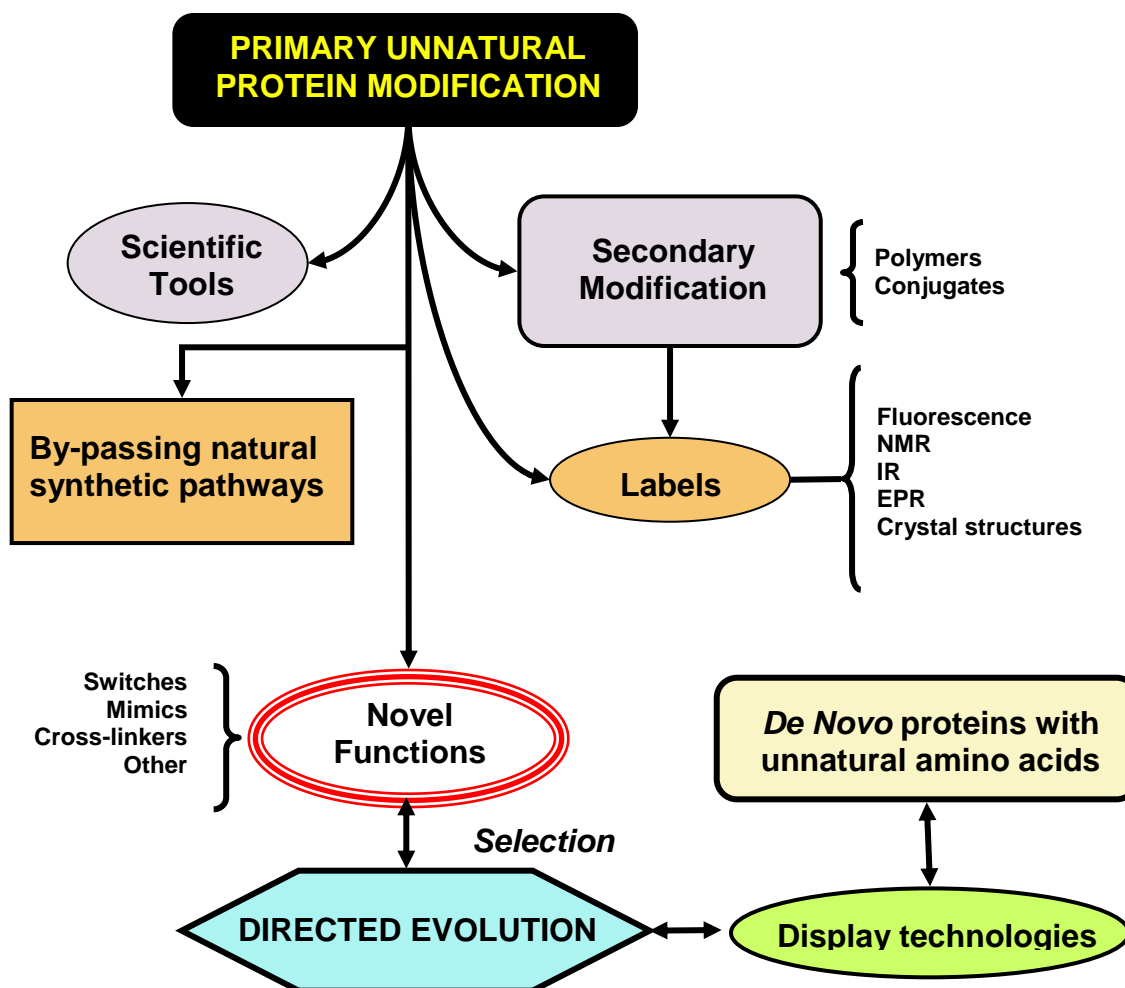
PRIMARY UNNATURAL PROTEIN MODIFICATION

Scientific Tools

Secondary Modification — Polymers, Conjugates

By-passing natural synthetic pathways

Labels — Fluorescence, NMR, IR, EPR, Crystal structures

Switches, Mimics, Cross-linkers, Other — **Novel Functions**

De Novo **proteins with unnatural amino acids**

Selection

DIRECTED EVOLUTION

Display technologies

Fig. 5.Nc

Utility of unnatural amino acids incorporated into proteins at predetermined sites through genetic code reassignments. 'Primary' unnatural modifications refer to unnatural amino acid incorporations which are directly functional, whether as labels for various spectroscopic or structural studies, or for directed evolutionary searches for novel diverse novel functions. 'Secondary' modification refers to the generation of protein conjugates at specifically predetermined sites, by means of an unnatural amino acid insertion which serves as a suitable reactive group for subsequent chemical adduct formation. *De novo* protein design can likewise make use of unnatural amino acids in a vast number of ways in principle. Abbreviations within 'Labels' for types of spectroscopy:

NMR, Nuclear Magnetic Resonance; IR, Infra-Red; EPR, Electron paramagnetic Resonance.

Unnatural amino acids which function as labels for infra-red and spin resonance (electron paramagnetic resonance) have also been incorporated into proteins by codon reassignments [5] (Fig. 5.Nc). For structural determinations by the important technique of nuclear magnetic resonance, labeling of proteins with amino acids bearing a fluorine-modified side chain is a useful approach, and this too has been achieved the 'unnatural' codon way [19]. Fluorinated variants of natural amino acids are also usually compatible with retention of native secondary structures [20], highly stable folds, and specific protein-protein interactions [21]. For structural studies using X-ray diffraction by protein crystals, co-incorporation of heavy atoms is a solution to the 'phasing problem', and for this purpose heavy halide ions (bromide, iodide) can be used [22]. Placement of such heavy atoms by precise incorporation of unnaturally-modified amino acids has many advantages for such purposes. Iodinated amino acids have been duly incorporated into unnatural genetic codes [23,24] and shown to assist crystal structural analyses [24].

An interesting application of the assignment of codons to unnatural amino acids is the enabling of ribosomal protein syntheses to perform what are normally post-translational enzymatic modifications. Consider an amino acid side chain R which is naturally modified *after* protein synthesis in certain proteins to RΞ, where the 'Ξ' modification can cover a wide range of different chemical groups. This action is normally completed by specific enzymes, as with kinases transferring phosphate groups to serine or threonine residues. There is no reason in principle, though, why the RΞ modified amino acid could not be incorporated directly during ribosomal protein synthesis, provided of course the necessary codon assignment, tRNA, and aminoacyl-tRNA synthetase were all in order and capable of handling the chosen amino acid modification. The results in either arrangement will be the same, since the RΞ amino acid is present in the final

processed proteins in both cases. But the ribosomal incorporation of RΞ is still quite 'unnatural', even though RΞ itself is not. A very common post-translational modification is glycosylation, and ribosomal incorporation of glycosylated amino acids through unnatural genetic coding has been demonstrated [25]. This process could be used to make proteins with a normal glycosylation pattern in the absence of the appropriate modifying enzymes, or extended further into the 'fully unnatural' realm by using different sugar residues, or changing the sites of glycosidic modifications. Another case in point in this general area concerns modifications which are normally rendered during non-ribosomal peptide syntheses by specific enzymes. One type of such unconventional peptide alterations are *N*-substitutions on the α-amino group (Fig. 5.9 of *Searching for Molecular Solutions*), and here too ribosomal encoding of such unnatural amino acids has succeeded, albeit in an *in vitro* system [26].

A field ripe for future development is the engineering of enzyme capabilities by the incorporation of functionally novel unnatural amino acids at or near active sites ♥. This (as an extension of conventional directed evolution) might also overcome some of the deficiencies of catalytic antibodies (Chapter 7). Engineering and controlling enzymatic pathways requires sophisticated control mechanisms including precise molecular switches, which in some cases are amenable to design using specific unnatural amino acid incorporation. Residues with side chain groups which are isomerizable or cleavable under defined conditions can be used in this regard. This issue of altering unnatural amino acids subsequent to their ribosomal incorporation leads us to the other major category of Fig. 5.Nc, or 'secondary' modifications. Chemical groups not found in natural proteins but useful for derivatization have been incorporated as unnatural genetically-encoded amino acids (including those bearing keto or azido groups [5]). Many of the labeling applications already referred to can also be approached

---

♥ There is precedent for the improvement of enzyme folding and stability by the incorporation of unnatural amino acids through synthetic means [27].

by this route (Fig. 5.Nc), which though indirect can increase the range of chemical options available for a single unnaturally-modified protein. Of all the chemical conjugates which might be produced at specific protein sites by such means, proteins modified with simple polymers may become the most commercially significant. Conjugation of proteins with polyethylene glycol (PEG) has been shown to improve their therapeutic efficacies [28], but heterogeneity of conventional adducts has been a problem [29]. One way around this is to target PEG derivatization ('PEGylation') to one or more defined unnaturally-incorporated reactive groups [30], although other options exist [28,29,31].

Section 16:   ***Specialized Ribosomes***

Cited on p. 186 of *Searching for Molecular Solutions*

Unnatural Factories

Whether it is carried out *in vitro* or *in vivo*, the biological synthesis of proteins requires ribosomes, which are remarkable macromolecular, multi-subunit, ribonucleoprotein processors of the information carried by mRNA. Fine-tuned by evolution over eons, the 'job' of the ribosome is fundamental to living systems whose effector and structural molecules are protein-based. Accordingly, ribosomes have high levels of sophistication in the efficient handling of the protein informational template molecules (mRNAs), the adaptor molecules between the genetic code and amino acids (tRNAs), the catalytic formation of peptide bonds, and the shepherding of nascent polypeptide chains. Ribosomes and associated systems also have means for detecting and dealing with problems during synthesis, such as translational stalling or defective mRNA templates. All of this, of course, has been evolutionarily honed for natural protein synthesis using the familiar set of building blocks, so there is no reason why it should be expected in advance that unnatural amino acids will necessarily be processed with equal efficiency. The experimental observation that a great many such non-natural 'raw materials' are in fact usable can be regarded as a bonus feature of biosynthetic protein production. But since protein creation through the agency of ribosomes is clearly not designed for the convenience of human protein engineers, at the same time it might be expected that certain limitations on the tolerance for unnatural tinkering will emerge.

It was noted in *Searching for Molecular Solutions* (Chapter 5) that β-amino acids and other amino acid alterations which change the natural α-carbon backbone have generally proved ineffective for insertion into natural protein sequences.

Although charging of tRNAs *in vitro* has very wide scope ♥, for polypeptide synthesis to proceed, the aminoacyl-tRNA must still be sterically compatible with its ribosomal interaction site and protein elongation factors. If natural ribosomes stubbornly refuse to cooperate with the translation of certain unnatural amino acids, would it not be then a logical step to redesign the ribosomes themselves? Before looking further at this seemingly audacious proposal, let us first briefly tour some of the salient features of natural ribosomes and their biosynthetic wizardry.

All proteins are made biologically by ribosomes translating mRNA molecules ♣, and it has long been known that ribosomes are ribonucleoprotein complexes of two major subunits ('large' and 'small' ♠). Each subunit itself is a complex of RNA molecules and many different proteins. Since the late 1990s, there has a been dramatic increase in our knowledge of ribosomal structures and functions, largely flowing from solving the crystal structures of prokaryotic ribosomes [36-40] (see Fig. 5.Nd). Three major functional sites pertaining to protein synthesis on the ribosome can be distinguished, based on their interactions with tRNA during the protein synthesis cycle (Fig. 5.Nd). Aminoacylated tRNA with the correct anticodon enters at the ribosomal A site where a complementary mRNA codon is

---

♥ *In vivo* charging of tRNAs also is limited by steric (size and shape) considerations of the amino acid binding sites of aminoacyl-tRNA synthetases; but chemical charging *in vitro* has no such restrictions.

♣ Note that this is *not* stating that all biological peptide bond formation is mediated by ribosomes. Many examples of both prokaryotic and eukaryotic non-ribosomal peptides are known [32,33], directed by non-ribosomal peptide synthetases [34]. But any biological polypeptide long enough to deserve to be called a protein will nevertheless be ribosome-derived.

♠ These are named by their centrifugal sedimentation characteristics ('S' or Svedberg units). Bacterial ribosomes are composed of large (50S) and small (30S) subunits, which associate to form the overall 70S ribosome. The general composition of eukaryotic ribosomes is broadly similar, although the major subunit sizes differ [35]. The large prokaryotic subunits are composed of two RNA strands (5S and 23S) and 36 proteins (L1-L36), and the small subunit has one RNA strand (16S) and 21 proteins (S1-S21) [36].

presented, and the growing peptide chain is transferred onto the amino group of the charged peptide for the entered tRNA. The peptidyl-tRNA is then translocated to the P site, and the deacylated tRNA formerly at the P site is translocated to the E (exit) site for release. The ribosome moves one codon in a 3' direction down the mRNA, and a new aminoacyl-tRNA then enters the A site [41]. Both ribosomal subunits contribute to the A, P, and E sites (Fig. 5.Nd), but catalysis of the peptidyl transferase reaction central to peptide elongation is restricted to the large subunit [41]. The physical separation of the initial substrate recognition and catalytic sites on ribosomes uniquely distinguishes them from other biosynthetic catalysts [42].
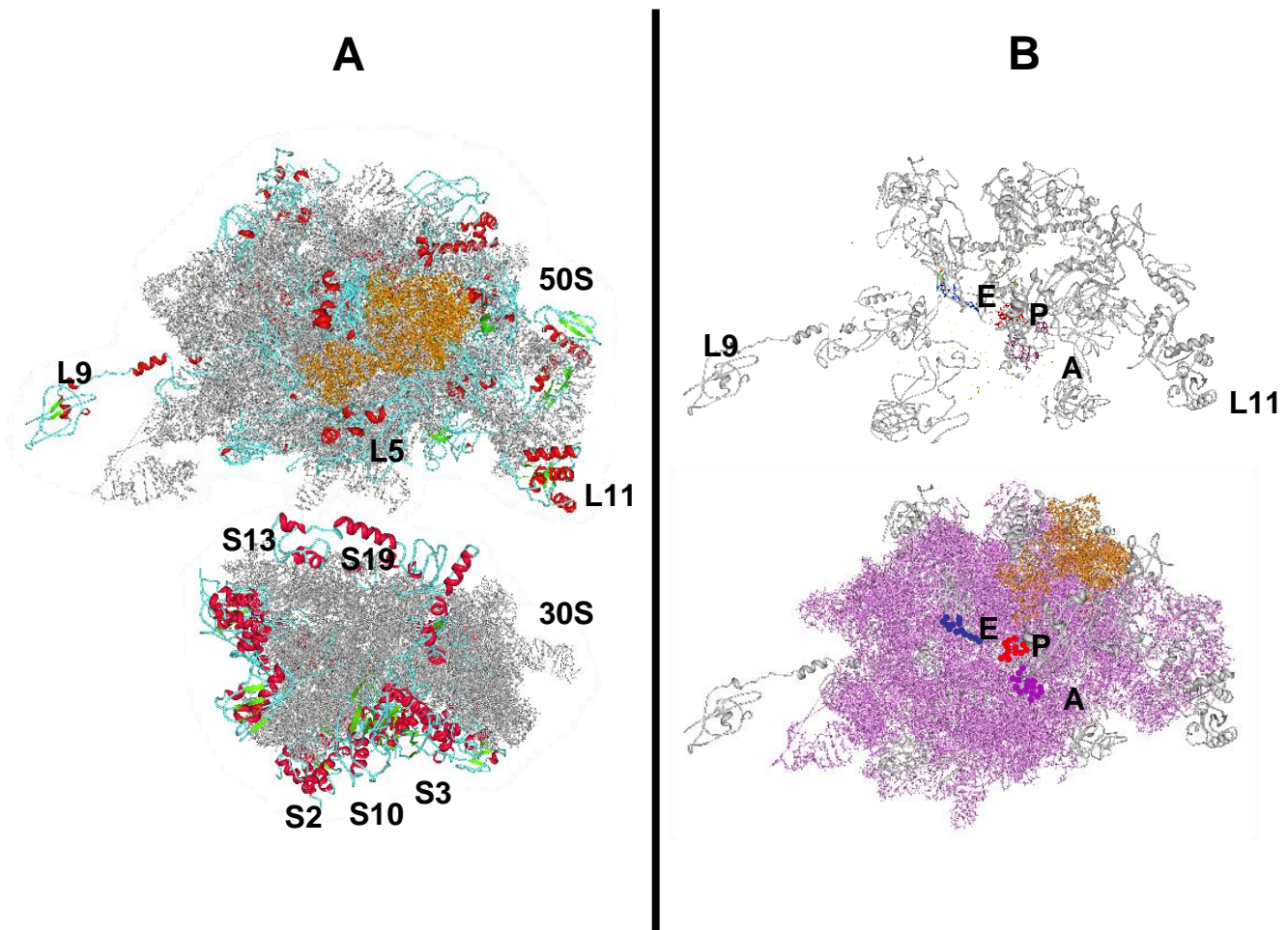

Fig. 5.Nd

Fig. 5. Nd. A, Structures of large (50S) and small (30S) ribosomal subunits of *E. coli* shown separated from their normal alignments. Visible proteins shown with α-helices red, β-strands green, turns blue, and several large and small subunit protein designations also are indicated. In large subunit, 23S RNA gray; 5S orange. In small subunit, 16S RNA gray. **B**, Top, large subunit showing proteins only (gray) except for key RNA residues at the E (exit; blue), P (peptidyl; red) and A (aminoacyl; dark purple) sites; bottom, same with 23S (light purple) and 5S (orange) RNAs superimposed. Source: <u>Protein Data Bank</u> [43] <u>2AW4</u> (50S) and <u>2AVY</u> (30S) [36]. Images generated with Protein Workshop [44].

One of the most interesting and biologically important findings to emerge from ribosome detailed structural studies has been the proof that the catalytic center for peptide bond synthesis is entirely derived from 23S ribosomal RNA. In other words, the ribosome is a giant RNA enzyme, or ribozyme [45-47]. Simply by measuring the physical distance between the site of peptidyl transfer and the nearest protein group ♥, it was evident that no protein-based catalytic mechanism was feasible [46]. At the same time, this is not to suggest that ribosomal proteins are completely irrelevant in this regard, but rather are likely to have an indirect role for the correct positioning of tRNAs than through direct catalytic participation [48]. Also, additional non-ribosomal proteins are essential for guiding the correct tRNAs to the A site, and for the molecular translocations necessary for each synthetic cycle (proteins EF-Tu and EF-G in prokaryotes, respectively [46]). Nevertheless, the finding of a universal fundamental role for ribozymes in one of the most essential and basic biological functions has profound ramifications for the 'RNA World' view of life's chemical evolution, and indeed the origin of life itself (as also noted in *Searching for Molecular Solutions*; Chapter 2). A logical extension of this work is to check for peptidyl transferase activity mediated by the

---

♥ 18 Angstrom units (Å) or 1.8×10^{-9} meters (1.8 nanometers), a significant separation distance in molecular terms.

ribosomal 23S subunit in isolation. Although this has not been successful to date
[49], non-ribosomal RNA catalytic activity of this type has been observed [50,51].

Increasing knowledge of a molecule or molecular system greatly enhances the chances of usefully manipulating or improving it. And so we can return to the question of redesigning ribosomes for specialized expression purposes with unnatural amino acids, where natural ribosomes are found wanting. Whenever one is hoping to modify essential gene products *in vivo*, an immediate problem presents itself. Any modification to an essential gene must be compatible with its normal function, or the effects will manifested as a lethal phenotype, and this dictate is obviously applicable to ribosomal function as a whole. The obvious solution to this problem (and a basic evolutionary mechanism) is to perform the manipulations on a second gene copy (a gene duplicate in natural settings, as was also considered in Chapter 2). But even this can become problematic if the manipulated gene copies express a product with 'dominant-negative' properties which can 'poison' the normal operations of the host cell. For example, certain mutant ribosomal RNAs can preferentially promote the translation of specific proteins, with lethal effects on the bacterial host cell's proteomic balance [52,53]. Nevertheless, generation of ribosomes with altered functional RNAs has been frequently undertaken, usually for testing structure/function predictions [41,54,55]. P (peptidyl transferase) sites are a logical target where improvement in the incorporation of unnatural amino acids is concerned. Following randomization of several residues associated with the P site in 23S RNA, ribosomal mutants which were compatible with cell growth ♥ were identified, and a number of these in turn demonstrated enhanced utilization *in vitro* of suppressor tRNAs charged with D-amino acids [56].

The reconstitution of ribosomes from their components *in vitro* has long been studied [57,58]. This process in principle affords an escape from problems

---

♥ A useful feature in this regard is the observation that certain P site mutations confer a degree of resistance to the antibiotic chloramphenicol [55], which is a protein synthesis inhibitor.

associated with the expression of certain mutant ribosomal RNAs *in vivo*, although when ribosomal RNAs from some organisms are transcribed *in vitro*, assembly into functional ribosomes is very inefficient ♥. But the ability to fully exploit engineered ribosomes *in vivo* is highly desirable, for similar reasons of economy and yield as noted in Chapter 5 of *Searching for Molecular Solutions* for efforts towards expansions of the genetic code.

From this point of view (and analogously with tRNA / aminoacyl-tRNA synthetase pairs which are orthogonal to each other), it would be a great advantage to generate ribosomes and mRNA which were both mutated and mutually orthogonal in their functions. Achieving this is surprisingly simple in principle. It is the 30S small ribosomal subunit which mediates recognition of most *E. coli* mRNAs, and this occurs through base-pairing interactions between sequences at the 5' end of prokaryotic mRNA molecules ('Shine-Dalgarno' sequences) and the 3' end of 16S RNA within the small ribosomal subunit ♣ [63]. Accordingly, if one arbitrarily altered the Shine-Dalgarno sequence in a specific mRNA and made *corresponding* changes to the 16S ribosomal RNA (the 'anti-Shine-Dalgarno' sequence) to preserve complementarity, in theory an orthogonal mRNA / ribosomal pair would be generated. This was performed decades ago ♠, with the addition of a useful trick involving another change in 16S RNA unrelated to the anti-Shine-Dalgarno sequence [64]. Spectinomycin blocks protein synthesis on normal ribosomes, but a specific 16S RNA mutation confers resistance [65]. In this

---

♥ Reconstitution of *E. coli* ribosomes with 23S RNA transcribed *in vitro* works poorly owing to a requirement for additional modifications to a critical region of this RNA [59]. On the other hand, reconstitution of certain other prokaryotic ribosomes from *in vitro* transcribed RNAs proceeds reasonably efficiently without additional RNA modifications [60,61]. In March 2009, the lab of George Church (Harvard) announced success at artificial ribosome generation.

♣ Some 'leaderless' mRNAs in bacteria can initiation translation on ribosomes through different mechanisms [62]

♠ For this kind of experiment, it is only necessary to express the mutant ribosomal RNA from an introduced plasmid vector; the expressed RNA will assemble with its appropriate ribosomal proteins expressed by the host.

manner, all host protein synthesis except for that occurring on mutant ribosomes can be shut down by spectinomycin treatment. The translational specificity of the ribosomal mutants (bearing combined altered anti-Shine-Dalgarno sequences plus the spectinomycin-resistance mutation) for corresponding mutant mRNA then becomes distinguishable from the background noise of host protein synthesis. The results provided by such means indicated that initial 'specialized ribosome / mRNA' pairs were not truly orthogonal, and indeed were lethal to the host cell if expressed constitutively [53]. Though useful for studying ribosomal RNA mutants [66], the 'orthogonality' of such prototype mutant ribosome / mRNA (Shine-Dalgarno / anti-Shine-Dalgarno) pairs needed considerable improvement to be useful for routine expression of unnatural amino acids.

And as in so many other cases we have considered, directed evolution again has come to the rescue. But this would appear at first glance to be a hard nut to crack from the combinatorics involved, as it is necessary to deal with the combination of pairs from two separate libraries of variants (ribosomal and mRNAs). To satisfy the demands of an ideal arrangement, each must be truly dedicated to each other and ignore normal host systems. The orthogonal mRNA must be translated only on the mutant ribosomes, and the latter must conversely fail to process host mRNAs, selectively translating only their cognate orthogonal mRNAs. This ideal exists among a larger set of cross-reactive possibilities, such as where mutant ribosomes are specific for mutant mRNAs, but not vice versa [67]. As we have seen from the derivation of orthogonal tRNAs and aminoacyl-tRNA synthetases, the secret of success is a combination of knowing where to target mutations (based on pre-existing information), and powerful positive and negative selections.

From libraries of variants of both an mRNA (spanning the Shine-Dalgarno sequence) and ribosomal 16S RNA (in the vicinity of the anti-Shine-Dalgarno sequence) orthogonal ribosome / mRNA pairs might exist, if a suitable selection process can be devised for their isolation. In fact, a novel fusion protein

conferring both a positively and negatively selectable phenotype ♥ has been indeed successfully used for this purpose. In this case, a library of mRNAs (where each member encodes this 'double-sieve' protein but varies at the Shine-Dalgarno sequence) was initially subjected to a negative selection. Any library members which can be expressed on normal ribosomes are removed by this selection process (depicted in Fig. 5.Ne), and the resulting sub-library can then be expressed along with the 16S ribosomal RNA anti-Shine-Dalgarno sequence library. Subjecting cells bearing all combinations of these libraries to a positive selection then allows an orthogonal mRNA / ribosomal pair to be derived (Fig. 5.Ne). Different sets of orthogonal ribosome / mRNA pairs with overlapping and predictable specificities can be generated through the same kind of selective processes, and the potential exploitation of these as molecular switches has also been noted [67].

The development of truly orthogonal pairs of mRNAs and ribosomes sets the stage for further evolution of ribosomes towards the acceptance of unnatural amino acids which are not tolerated by natural ribosomes, and to improve the efficiency of ribosomes towards specific tRNA molecules widely used for insertion of unnatural amino acids during protein synthesis. As noted earlier, amber-suppressor tRNAs orthogonally charged with unnatural amino acids have been used to expand the genetic code *in vivo*. Although demonstrably successful, this expression is still limited in its efficiency in comparison with normal circumstances. One problem identified as responsible (at least in part) for such inefficiencies is competition for the reassigned amber (UAG) codon with a protein release factor, RF-1. This release factor competition also occurs with *in*

♥ This study [67] developed a fusion gene with the resistance gene for the antibiotic chloramphenicol (chloramphenicol acetyl transferase; positive selection) and uracil phosphoribosyltransferase (UPRT). When UPRT is present, the compound 5-fluorouracil is converted into a product which inhibits the essential enzyme thymidylate synthetase. Consequently, 5-fluorouracil can be used to negatively select cells expressing UPRT, while cells not expressing UPRT continue to grow.

vitro expression systems, and it has been shown that partial *in vitro* inactivation of release factor improves incorporation efficiency of unnatural amino acids [68]. This cannot be done so simply *in vivo*, as release factors are essential cellular proteins, as might be expected. But since the initial delivery of tRNAs to ribosomes is via the A site (Fig. 5.Nd), altering the A site might at least partially overcome this problem. Using an orthogonal mRNA / ribosome system, libraries of A site mutants were accordingly used to select variants with significantly improved utilization of suppressor tRNAs [69].
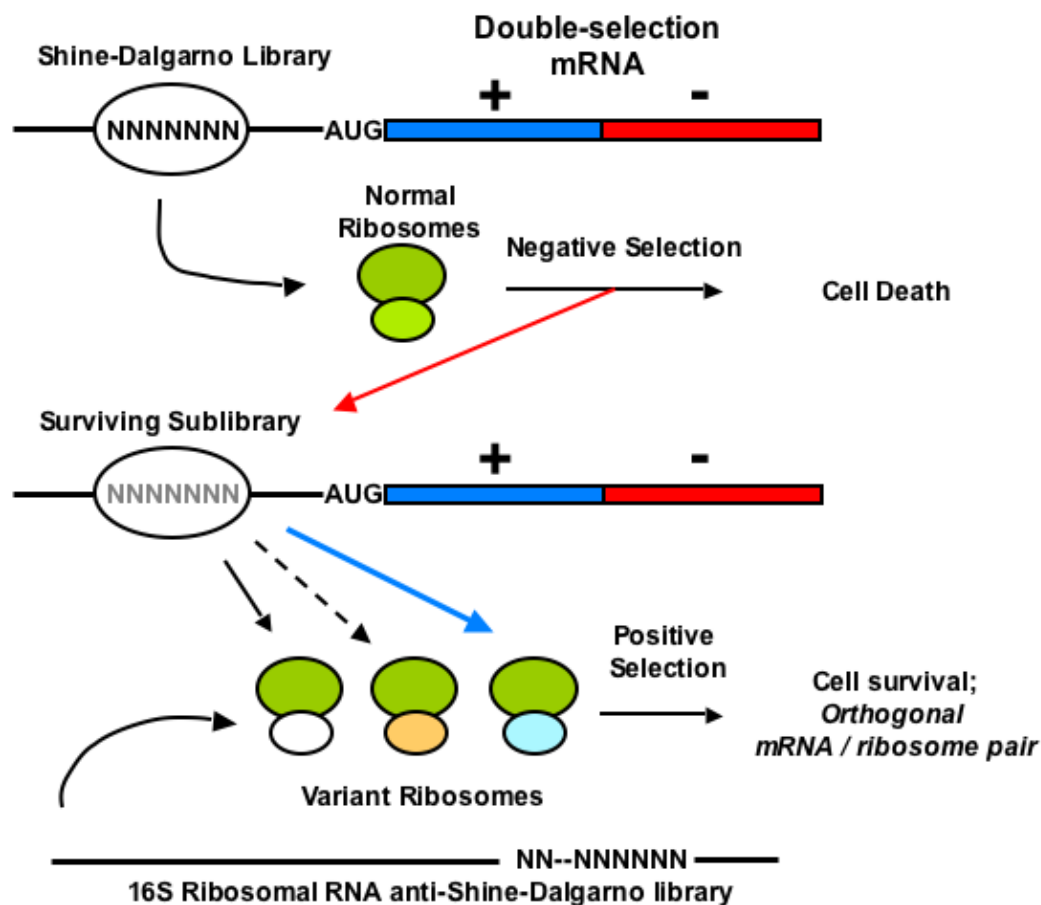

Fig. 5.Ne

Fig. 5.Ne. Process for selection of orthogonal mRNA / ribosome pairs with a fusion gene for both positive and negative selections.

Since the generic sequence complementarity requirements for orthogonality between 16S RNA and mRNA molecules are easy to demarcate in themselves (orthogonal pair mutual recognition with neither interacting with host molecules), rational computational design of relevant sets of specific sequences has also been applied. This resulted in considerable success, yielding a number of identified pairs with high orthogonality and without significant toxicity [70]. Such analyses can in principle be applied to a wide range of different organisms. Many more developments with the engineering of orthogonal mRNA / ribosome systems can be expected, which will have considerable impact upon both direct biotechnological applications and the continual fine-tuning of our understanding of ribosomal functional mechanisms. A notable recent innovation is the use of an orthogonal ribosome / mRNA system to develop transcription-translation loops which would not be possible using normal host systems [71].

Before moving on from this topic, though, we should make note of a non-ribosomal factor which may also limit the utilization of unnatural amino acids. The elongation factor Tu (EF-Tu) is responsible for delivering charged tRNA molecules to the ribosomal A site (as briefly noted above), and this protein has an intriguing recognition mode towards aminoacyl-tRNA molecules. Misacylation of the same tRNA with a range of different amino acids resulted in binding by EF-Tu over a wide range of affinities (either greater or less than that shown towards correctly aminoacylated tRNAs) [72]. Conversely, a variety of different tRNAs misacylated with the same amino acid also showed a wide range of EF-Tu binding affinities [73,74].

These results suggested that EF-Tu recognition of natural aminoacylated tRNAs is dependent on motifs present in both the tRNA and amino acid moieties.

Variation in EF-Tu binding affinities towards different natural amino acids charged onto their cognate tRNAs can be 'off-set' by other changes within the same tRNA (or 'thermodynamically compensated' [72,73]) such that biologically useful binding of all aminoacyl-tRNAs by a single EF-Tu protein is maintained ♥. Assembling such data allows accurate predictions of the free energy of binding (a measure of affinity) for EF-Tu towards misacylated tRNAs [75]. Moreover, if EF-Tu binding for a misacylated tRNA (such as a suppressor tRNA bearing an unnatural amino acid) is below a certain threshold, the charged tRNA may fail to be translationally active [75]. Future improvements in the use of unnatural amino acids may therefore need to take the role of EF-Tu into account, possibly by the directed evolution of EF-Tu variants orthogonal towards specific unnatural aminoacylated tRNAs. Variants of EF-Tu in some organisms which recognize corresponding unusual tRNA variants may provide useful information in this regard [76].

References:

1. Nagendrappa, G. Benzene and its Isomers. How many structures can we draw for C6H6? *Resonance*, 74-78 (2001).

2. Caldwell, J. & Wainer, I. W. Stereochemistry: definitions and a note on nomenclature. *Hum Psychopharmacol* **16**, S105-S107 (2001).

3. Morrison, R. T. & Boyd, R. N. *Organic Chemistry* (Addison Wesley Longman, 1999).

4. Blattner, F. R. et al. The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-74 (1997).

5. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu Rev Biophys Biomol Struct* **35**, 225-49 (2006).

6. Ibba, M. & Soll, D. Aminoacyl-tRNA synthesis. *Annu Rev Biochem* **69**, 617-50 (2000).

7. Steer, B. A. & Schimmel, P. Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *J Biol Chem* **274**, 35601-6 (1999).

♥ Significant variation in EF-Tu binding affinities towards the range of correctly aminoacylated tRNAs is seen, but less than that observed with misacylated tRNAs [72].

8.   Santoro, S. W. & Schultz, P. G. Directed evolution of the substrate specificities of a site-specific recombinase and an aminoacyl-tRNA synthetase using fluorescence-activated cell sorting (FACS). *Methods Mol Biol* **230**, 291-312 (2003).

9.   Nureki, O. et al. Structural basis for amino acid and tRNA recognition by class I aminoacyl-tRNA synthetases. *Cold Spring Harb Symp Quant Biol* **66**, 167-73 (2001).

10.   Xie, J. & Schultz, P. G. A chemical toolkit for proteins--an expanded genetic code. *Nat Rev Mol Cell Biol* **7**, 775-82 (2006).

11.   Longstaff, D. G. et al. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc Natl Acad Sci U S A* **104**, 1021-6 (2007).

12.   Mehl, R. A. et al. Generation of a bacterium with a 21 amino acid genetic code. *J Am Chem Soc* **125**, 935-9 (2003).

13.   Hutter, H. Fluorescent reporter methods. *Methods Mol Biol* **351**, 155-73 (2006).

14.   Shaner, N. C., Patterson, G. H. & Davidson, M. W. Advances in fluorescent protein technology. *J Cell Sci* **120**, 4247-60 (2007).

15.   Summerer, D. et al. A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* **103**, 9785-9 (2006).

16.   Wang, L., Xie, J., Deniz, A. A. & Schultz, P. G. Unnatural amino acid mutagenesis of green fluorescent protein. *J Org Chem* **68**, 174-6 (2003).

17.   Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid. *J Am Chem Soc* **128**, 8738-9 (2006).

18.   Zhang, Z. et al. Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc Natl Acad Sci U S A* **101**, 8882-7 (2004).

19.   Jackson, J. C., Hammill, J. T. & Mehl, R. A. Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J Am Chem Soc* **129**, 1160-6 (2007).

20.   Link, A. J. & Tirrell, D. A. Reassignment of sense codons in vivo. *Methods* **36**, 291-8 (2005).

21.   Yoder, N. C. & Kumar, K. Fluorinated amino acids in protein design and engineering. *Chem Soc Rev* **31**, 335-41 (2002).

22.   Dauter, M. & Dauter, Z. Phase determination using halide ions. *Methods Mol Biol* **364**, 149-58 (2007).

23.   Sakamoto, K. et al. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res* **30**, 4692-9 (2002).

24.   Xie, J. et al. The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotechnol* **22**, 1297-301 (2004).

25. Zhang, Z. et al. A new strategy for the synthesis of glycoproteins. *Science* **303**, 371-3 (2004).

26. Kawakami, T., Murakami, H. & Suga, H. Exploration of incorporation of Nalpha-methylated amino acids into peptides by sense-suppression method. *Nucleic Acids Symp Ser (Oxf)*, 361-2 (2007).

27. Arnold, U. et al. Protein prosthesis: a nonnatural residue accelerates folding and increases stability. *J Am Chem Soc* **125**, 7500-1 (2003).

28. Brocchini, S. et al. Disulfide bridge based PEGylation of proteins. *Adv Drug Deliv Rev* **60**, 3-12 (2008).

29. Brocchini, S. et al. PEGylation of native disulfide bonds in proteins. *Nat Protoc* **1**, 2241-52 (2006).

30. Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M. & Schultz, P. G. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg Med Chem Lett* **14**, 5743-5 (2004).

31. Balan, S. et al. Site-specific PEGylation of protein disulfide bonds using a three-carbon bridge. *Bioconjug Chem* **18**, 61-76 (2007).

32. Welker, M. & von Dohren, H. Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiol Rev* **30**, 530-63 (2006).

33. Stack, D., Neville, C. & Doyle, S. Nonribosomal peptide synthesis in Aspergillus fumigatus and other fungi. *Microbiology* **153**, 1297-306 (2007).

34. Challis, G. L. & Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* **14**, 748-56 (2004).

35. Lewin, B. *Genes IX* (Jones and Bartlett, 2007).

36. Schuwirth, B. S. et al. Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310**, 827-34 (2005).

37. Ban, N. et al. A 9 A resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* **93**, 1105-15 (1998).

38. Wimberly, B. T. et al. Structure of the 30S ribosomal subunit. *Nature* **407**, 327-39 (2000).

39. Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-42 (2006).

40. Korostelev, A. & Noller, H. F. The ribosome in focus: new structures bring new insights. *Trends Biochem Sci* **32**, 434-41 (2007).

41. Moore, P. B. & Steitz, T. A. The structural basis of large ribosomal subunit function. *Annu Rev Biochem* **72**, 813-50 (2003).

42. Tan, Z., Forster, A. C., Blacklow, S. C. & Cornish, V. W. Amino acid backbone specificity of the Escherichia coli translation machinery. *J Am Chem Soc* **126**, 12752-3 (2004).

43.   Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980 (2003).

44.   Moreland, J. L., Gramada, A., Buzko, O. V., Zhang, Q. & Bourne, P. E. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **6**, 21 (2005).

45.   Moore, P. B. & Steitz, T. A. The involvement of RNA in ribosome function. *Nature* **418**, 229-35 (2002).

46.   Steitz, T. A. & Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* **28**, 411-8 (2003).

47.   Beringer, M. & Rodnina, M. V. The ribosomal peptidyl transferase. *Mol Cell* **26**, 311-21 (2007).

48.   Maguire, B. A., Beniaminov, A. D., Ramu, H., Mankin, A. S. & Zimmermann, R. A. A protein component at the heart of an RNA machine: the importance of protein l27 for the function of the bacterial ribosome. *Mol Cell* **20**, 427-35 (2005).

49.   Anderson, R. M., Kwon, M. & Strobel, S. A. Toward ribosomal RNA catalytic activity in the absence of protein. *J Mol Evol* **64**, 472-83 (2007).

50.   Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-8 (2002).

51.   Sun, L., Cui, Z., Gottlieb, R. L. & Zhang, B. A selected ribozyme catalyzing diverse dipeptide synthesis. *Chem Biol* **9**, 619-28 (2002).

52.   Jacob, W. F., Santer, M. & Dahlberg, A. E. A single base change in the Shine-Dalgarno region of 16S rRNA of Escherichia coli affects translation of many proteins. *Proc Natl Acad Sci U S A* **84**, 4757-61 (1987).

53.   Lee, K., Holland-Staley, C. A. & Cunningham, P. R. Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *Rna* **2**, 1270-85 (1996).

54.   Polacek, N., Gaynor, M., Yassin, A. & Mankin, A. S. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* **411**, 498-501 (2001).

55.   Thompson, J. et al. Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc Natl Acad Sci U S A* **98**, 9002-7 (2001).

56.   Dedkova, L. M., Fahmi, N. E., Golovine, S. Y. & Hecht, S. M. Enhanced D-amino acid incorporation into protein by modified ribosomes. *J Am Chem Soc* **125**, 6616-7 (2003).

57.   Nomura, M. Assembly of bacterial ribosomes. *Science* **179**, 864-73 (1973).

58.   Nomura, M. Reflections on the days of ribosome reconstitution research. *Trends Biochem Sci* **22**, 275-9 (1997).

59.   Green, R. & Noller, H. F. In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. *Rna* **2**, 1011-21 (1996).

60.   Green, R. & Noller, H. F. Reconstitution of functional 50S ribosomes from in vitro transcripts of Bacillus stearothermophilus 23S rRNA. *Biochemistry* **38**, 1772-9 (1999).

61.   Khaitovich, P., Tenson, T., Kloss, P. & Mankin, A. S. Reconstitution of functionally active Thermus aquaticus large ribosomal subunits with in vitro-transcribed rRNA. *Biochemistry* **38**, 1780-8 (1999).

62.   Moll, I., Hirokawa, G., Kiel, M. C., Kaji, A. & Blasi, U. Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res* **32**, 3354-63 (2004).

63.   Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233-41 (2001).

64.   Hui, A. & de Boer, H. A. Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in Escherichia coli. *Proc Natl Acad Sci U S A* **84**, 4762-6 (1987).

65.   Sigmund, C. D., Ettayebi, M. & Morgan, E. A. Antibiotic resistance mutations in 16S and 23S ribosomal RNA genes of Escherichia coli. *Nucleic Acids Res* **12**, 4653-63 (1984).

66.   Hui, A. S., Eaton, D. H. & de Boer, H. A. Mutagenesis at the mRNA decoding site in the 16S ribosomal RNA using the specialized ribosome system in Escherichia coli. *Embo J* **7**, 4383-8 (1988).

67.   Rackham, O. & Chin, J. W. A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol* **1**, 159-66 (2005).

68.   Short, G. F., 3rd, Golovine, S. Y. & Hecht, S. M. Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry* **38**, 8808-19 (1999).

69.   Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* **25**, 770-7 (2007).

70.   Chubiz, L. M. & Rao, C. V. Computational design of orthogonal ribosomes. *Nucleic Acids Res* **36**, 4038-46 (2008).

71.   An, W. & Chin, J. W. Synthesis of orthogonal transcription-translation networks. *Proc Natl Acad Sci U S A* **106**, 8477-82 (2009).

72.   LaRiviere, F. J., Wolfson, A. D. & Uhlenbeck, O. C. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* **294**, 165-8 (2001).

73.   Asahara, H. & Uhlenbeck, O. C. The tRNA specificity of Thermus thermophilus EF-Tu. *Proc Natl Acad Sci U S A* **99**, 3499-504 (2002).

74. Dale, T., Sanderson, L. E. & Uhlenbeck, O. C. The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry* **43**, 6159-66 (2004).

75. Asahara, H. & Uhlenbeck, O. C. Predicting the binding affinities of misacylated tRNAs for Thermus thermophilus EF-Tu.GTP. *Biochemistry* **44**, 11254-61 (2005).

76. Ohtsuki, T. & Watanabe, Y. T-armless tRNAs and elongated elongation factor Tu. *IUBMB Life* **59**, 68-75 (2007).