Introduction of non-natural amino acids and bases

1) synthetic peptides (<50 amino acids)
2) native chemical ligation (allows the generation of larger proteins containing synthetic elements)

Thiol ester
- Made by solid-phase synthesis

-N-terminal Cys fragment

From resins that look like:

Intein ligation

Inteins in protein splicing:

thiol to amino-acyl migration

requires cysteine residue at site of ligation (can be removed chemically)

(Cys)
Large protein: 200 aas

3) Biosynthesis of proteins with unnatural amino acids
- 64 triplets code for 20 amino acids
- 3 triplets code for stop (nonsense codon)

Process:
1) "blank" codons (nonsense, four base, rare codons, condon/tRNA deleted E. coli genome, etc.)
2) Engineer a tRNA that is orthogonal to all other tRNAs
3) Evolve an aminoacyl-tRNA synthetase to uniquely recognize this tRNA
4) Evolve a synthetase to uniquely charge this tRNA with the 21st amino acid
5) Biosynthesize or transport amino acid (most amino acids are transported into bacteria as the monomer or LysX dipeptide)

YYYXXX
AAAUUU

Peptidyl transferase - broad substrate specificity
-tRNA

migration from A to P site
to add another amino acid you need a unique codon that codes for the new amino acid, as well as a tRNA specific for it.

-use the stop codon TAG (UAG) to encode unnatural amino acids

Usually:

"stop"
UAG

Instead use UAA as the stop codon, and UAG to code for unnatural amino acids

2) Build a tRNA that recognizes UAG

20 natural aaRS

Archae bacteria

If you import a tRNA-aaRS pair from archae into E. coli, they do not cross react with E.coli tRNAs or synthetases but are functional in translation

3) Evolve an aminoacyl-tRNA synthetase (aaRS) to uniquely recognize this tRNA and no endogenous E. coli tRNAs

Tyrosyl tRNA synthetase:

ATP + Tyr $\rightarrow$ AMP + $\text{NH}_2 + \text{PP}_i$

Nonsense sequence tRNA cannot be recognized by endogenous synthetases

identity element in every tRNA except for proline in E. coli

identity elements distinct from E. coli

If you import a tRNA-aaRS pair from archae into E. coli, they do not cross react with E.coli tRNAs or synthetases but are functional in translation
But there is a problem: When you change the anticodon to CUA, the identity uniqueness is lost, and in addition to the Tyr aaRS (archae) recognizing it, some E. coli aaRS also recognize and load the tRNA. The solution is to create a large library of tRNAs and use an in vitro selection scheme to identify an orthogonal one.

AUC

AA

UAG

barnase (RNAse)-a toxic gene that kills cells when non-orthogonal tRNAs are recognized by host aaRS (negative selection)

TAG at permissive site (accepts all 20 amino acids)

15-20 nucleotide mutations to N (A,G,C,T) -the only cells that live are ones in which the tRNA is not recognized by any E. coli aaRS or are non functional in translation, i.e.:

- no longer recognized by archaea Tyr aaRS
- no longer functional on the ribosome

Solution: Take all the tRNA winners and transform E. coli along with the archae Tyr aaRS.

β-lactamase allows growth on ampicillin - positive selection

RNA winners

β-lactamase recognizes Tyr aaRS

20 mutations

AUC

Take all the winners-load NAla or endogenous host amino acid

- NAla and 20 canonical amino acids

Want to substitute Tyr with NAla then:

Generate a structure based randomized library where the 6 active residues binding to tyrosine are randomized.

Permissive site for NAla

NAla

aaRS

library

AUC

ATP

Tyr

AMP

Tyr

H2N

O

Na

OH

Evolve a synthetase to uniquely charge this tRNA with the 21st amino acid and not endogenous host amino acids.

4)
To remove aaRS that recognize endogenous host amino acids take the winners:

-20 canonical amino acids, but no NAla

Anything remaining does not take any canonical amino acids so must be specific for NAla

**Labelling:**

- Nitrene (CH insertion affinity label)
- CH insertion

**Metal binding amino acids:**

**Photocaged amino acids:**

Using isoGTP, isoCTP, dTTP* (radio-labelled), dGTP, dCTP

Then using isoGTP, dCTP, dTTP*

Expanding the number of base pairs in DNA and RNA (Benner)

\[
\begin{align*}
A & \equiv T \\
G & \equiv C \\
X & \equiv Y
\end{align*}
\]

Stable, base pairs with itself, does not base pair with A, G, C or T

-3H bonds
-Orthogonal H-bonding patterns

Using isoGTP, isoCTP, dTTP* (radio-labelled), dGTP, dCTP

Then using isoGTP, dCTP, dTTP*

Tautomer of IsoG, has comparable stability and pairs with T

**Expanding the number of base pairs in DNA and RNA (Benner)**
Expanding the number of base pairs in DNA and RNA (DiPic)

To measure these look at Tm

<table>
<thead>
<tr>
<th></th>
<th>Tm(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiPic-Py</td>
<td>40.4</td>
</tr>
<tr>
<td>DiPic-A</td>
<td>36.1</td>
</tr>
<tr>
<td>DiPic (G,C,T)</td>
<td>&lt;30</td>
</tr>
<tr>
<td>Py (A,G,C,T)</td>
<td>&lt;30</td>
</tr>
</tbody>
</table>

Another way (that didn't work): use a tricycle and monocycle pair

longer tricycle  shorter