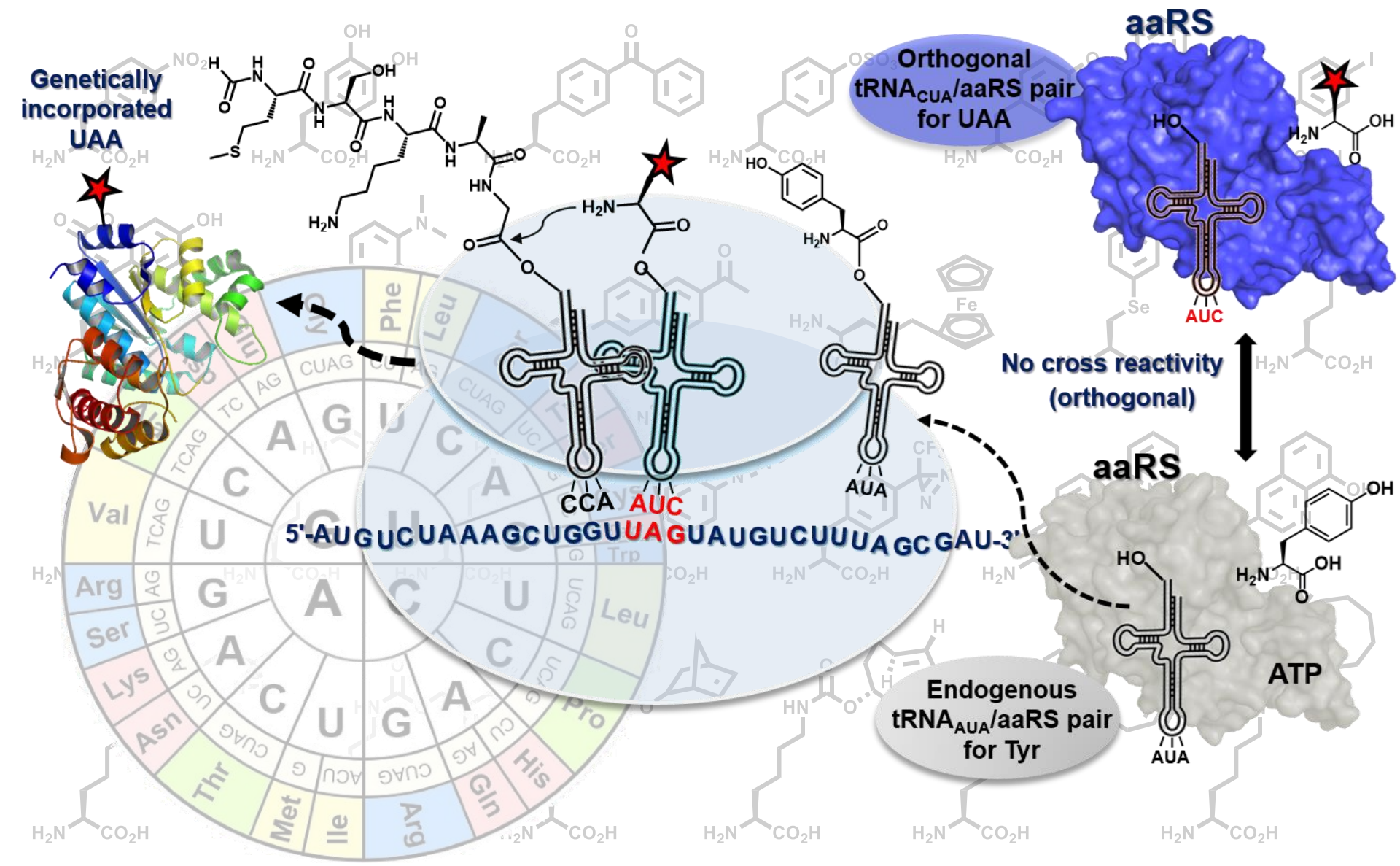
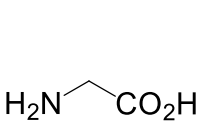
CHAPTER 9

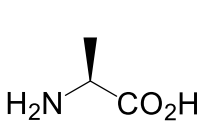
Genetic Code Expansion Technology



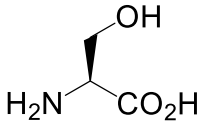
20 Canonical Amino Acids



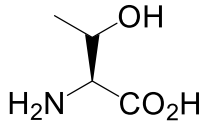
Gly, G



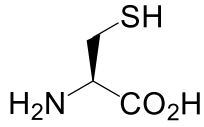
Ala, A



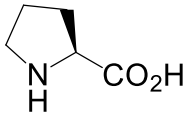
Ser, S



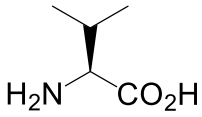
Thr, T



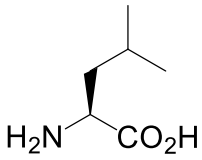
Cys, C



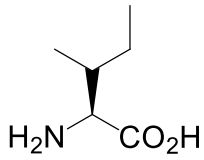
Pro, P



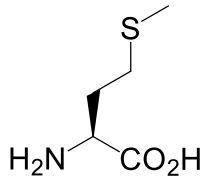
Val, V



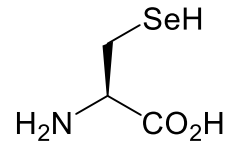
Leu, L



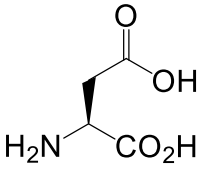
Ile, I



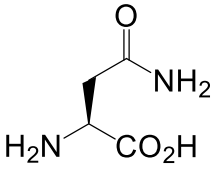
Met, M



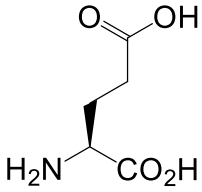
Selenocysteine
SeCys



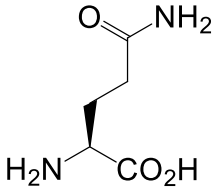
Asp, D



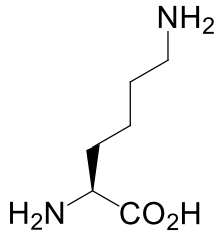
Asn, N



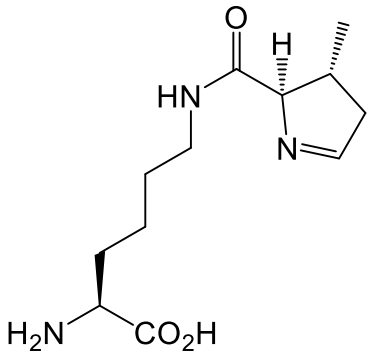
Glu, E



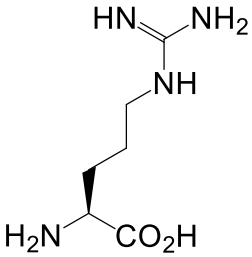
Gln, Q



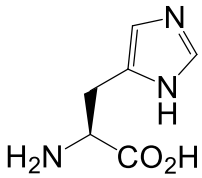
Lys, K



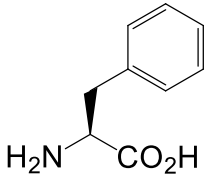
Pyrrolysine
Pyl



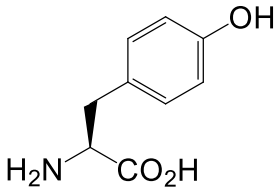
Arg, R



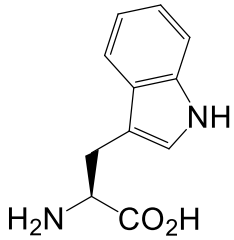
His, H



Phe, F



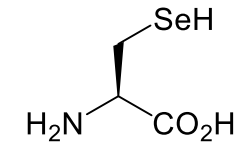
Tyr, Y



Trp, W

Selenocysteine

- Selenocysteine is the 21st proteinogenic amino acid existing in all three domains of life
- SeCys is present in several enzymes including glutathione peroxidases and formate dehydrogenases
- In 1976, the amino acid was first discovered in clostridial glycine reductase *PNAS* **1976**, *73*, 2659.
- SeCys in enzymes catalyzes oxido-reduction
  - SeCys has both a lower pK_a (5.47) and a lower reduction potential than cysteine
- SeCys is not coded for directly in the genetic code – no codon is available



Selenocysteine
SeCys

◆ tRNA for SeCys

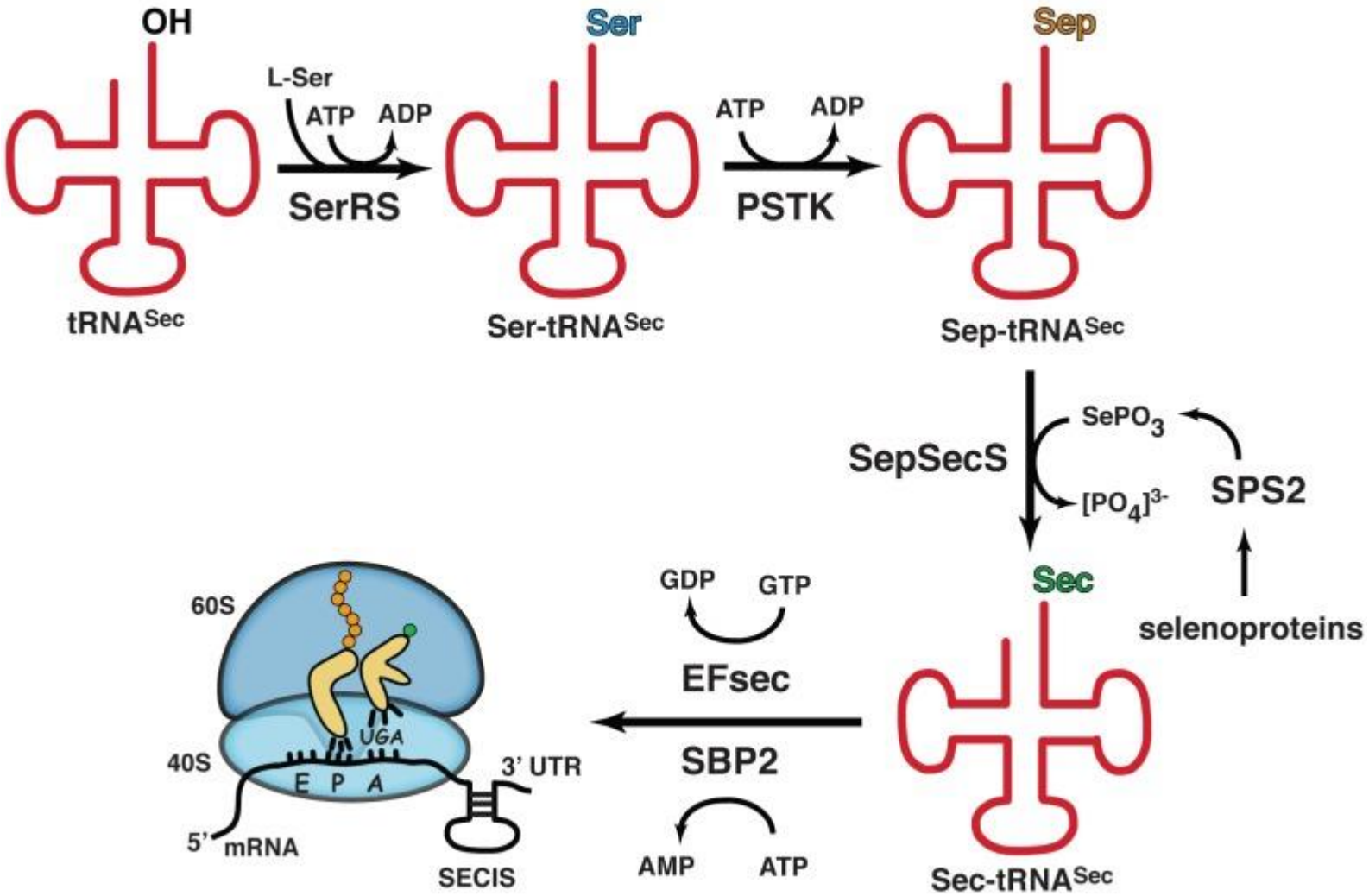
Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine

Walfred Leinfelder*, Eva Zehelein*, Marie-Andreé Mandrand-Berthelot† & August Böck*

Nature **1988**, *331*, 723.

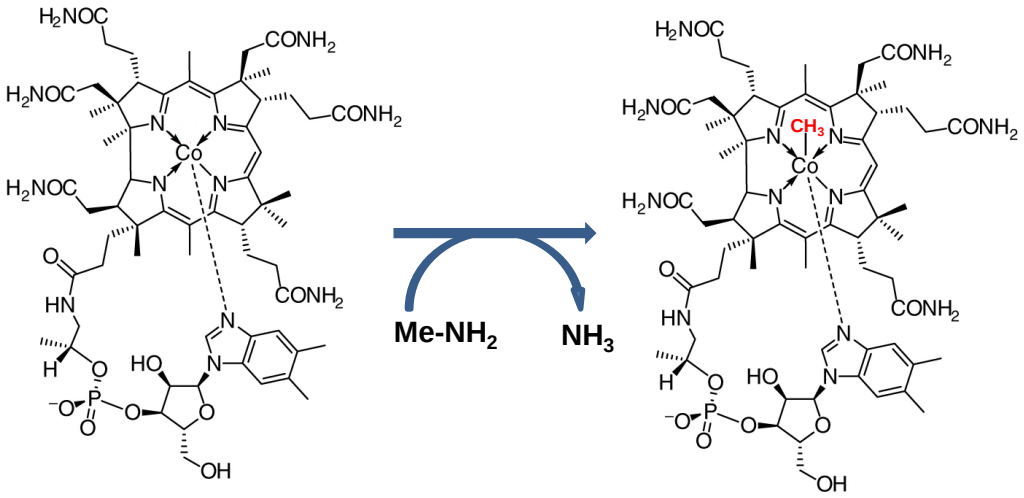
- Sequence analysis of the genes coding for two selenoproteins, formate dehydrogenase H from E. coli and glutathione peroxidase from mouse and man, demonstrated that *an inframe UGA opal nonsense codon directs the incorporation of selenocysteine*
- Recently, we identified four genes whose products are required for selenocysteine incorporation in E. coli
- We report here that one of these genes codes for a tRNA species with unique properties
- *It possesses an anticodon complementary to UGA* and deviates in several positions from sequences, until now, considered invariant in all tRNA species
- *This tRNA is aminoacylated with L-serine by the seryl-tRNA ligase* which also charges cognate $tRNA_{Ser}$. Selenocysteine, therefore, is synthesized from a serine residue bound to a natural suppressor tRNA which recognizes UGA.

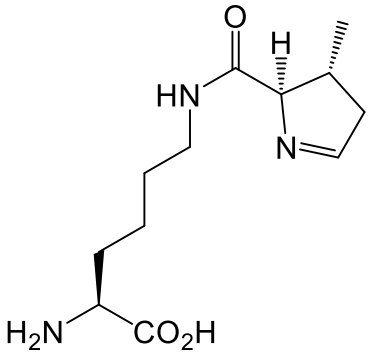
◆ Conversion of Ser into SeCys



Pyrrolysine

- Pyrrolysine (Pyl) an α-amino acid that is used in the biosynthesis of proteins in some methanogenic archaea and bacteria
- The extra pyrroline ring is incorporated into the active site of several methyltransferases
- Pyl is required to transfer a methyl group to the cofactor, corrinoid





Pyrrolysine
Pyl

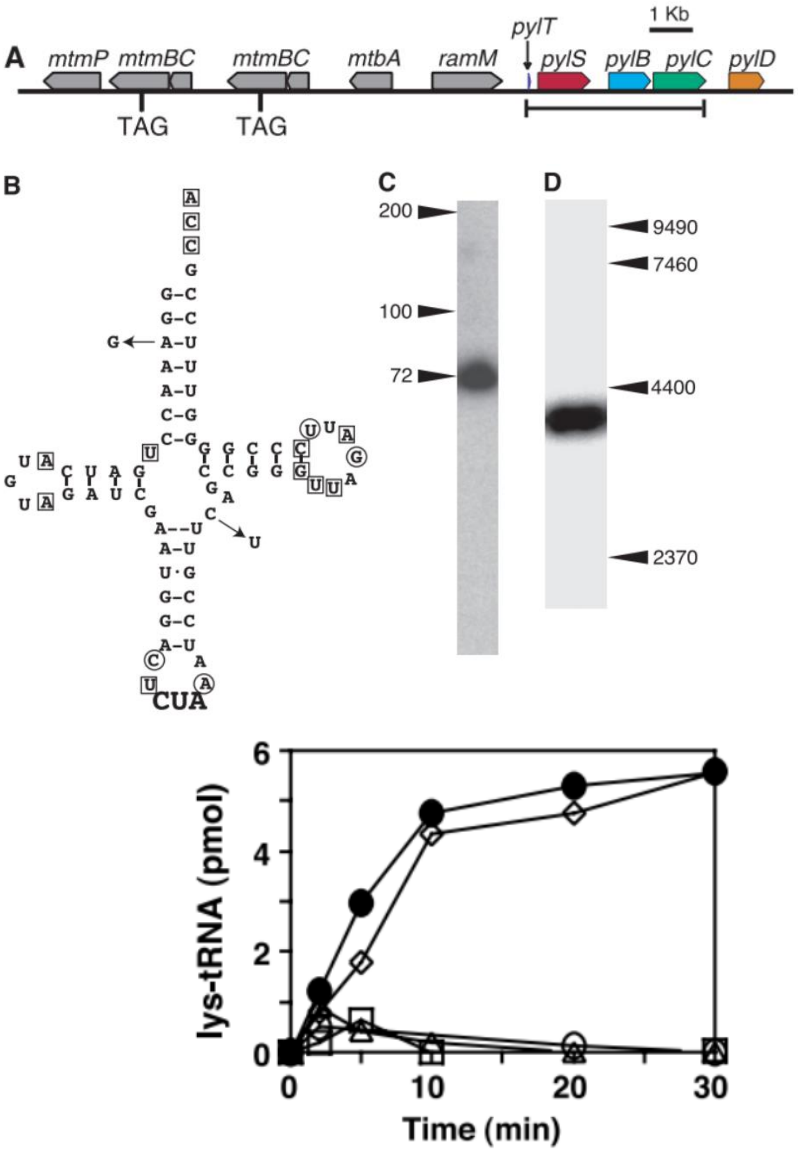
How Pyl is encoded

Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA

Gayathri Srinivasan, Carey M. James, Joseph A. Krzycki*

Science **2002**, *296*, 1459.

- Pyrrolysine is a lysine derivative encoded by the UAG codon in methylamine methyltransferase genes of *Methanosarcina barkeri*.
- Near a methyltransferase gene cluster is the pylT gene, which encodes an unusual transfer RNA (tRNA) with a CUA anticodon.
- The adjacent pylS gene encodes a class II aaRS that charges the pylT-derived tRNA with lysine but is not closely related to known lysyl-tRNA synthetases
- Homologs of pylS and pylT are found in a Gram-positive bacterium. Charging a $tRNA_{CUA}$ with lysine is a likely first step in translating UAG amber codons as pyrrolysine in certain methanogens.
- Our results indicate that pyrrolysine is the 22nd genetically encoded natural amino acid.



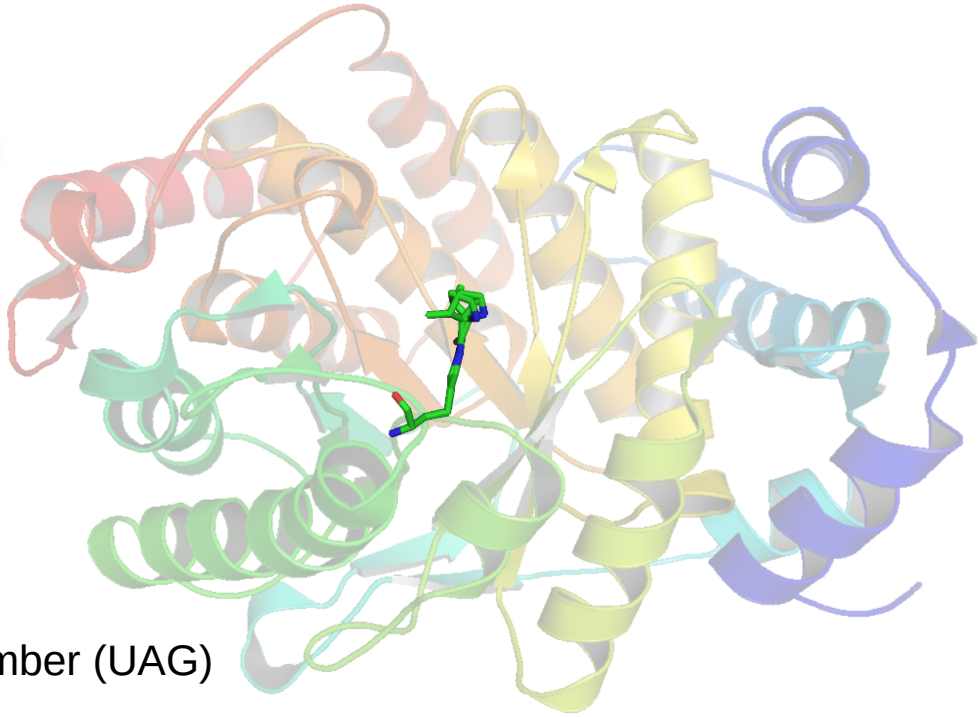
Acylation by LysRS to pylT.
At that time, Pyl had not been isolated or synthesized, and it was not available

How Pyl is encoded

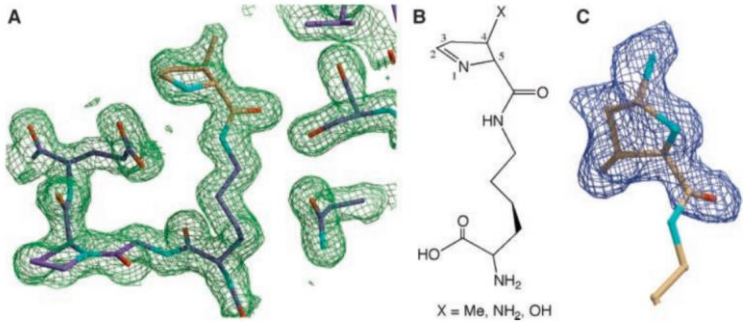
A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase

Bing Hao,[1] Weimin Gong,[1] Tsuneo K. Ferguson,[2]
Carey M. James,[2] Joseph A. Krzycki,[2]* Michael K. Chan[1]*

Science **2002**, *296*, 1462.



- Genes encoding methanogenic methylamine methyltransferases all contain an in-frame amber (UAG) codon that is read through during translation.
- We have identified the UAG-encoded residue in a 1.55 Å resolution structure of the Methanosarcina barkeri monomethylamine methyltransferase (MtmB).
- This structure reveals a homohexamer comprised of individual subunits with a TIM barrel fold.
- The electron density for the UAG-encoded residue is distinct from any of the 21 natural amino acids.
- Instead it appears consistent with a lysine in amide-linkage to (4R,5R)-4-substituted-pyrroline-5-carboxylate. We suggest that this amino acid be named L-pyrrolysine.



◆ How Pyl is encoded

Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vitro* and *in vivo*

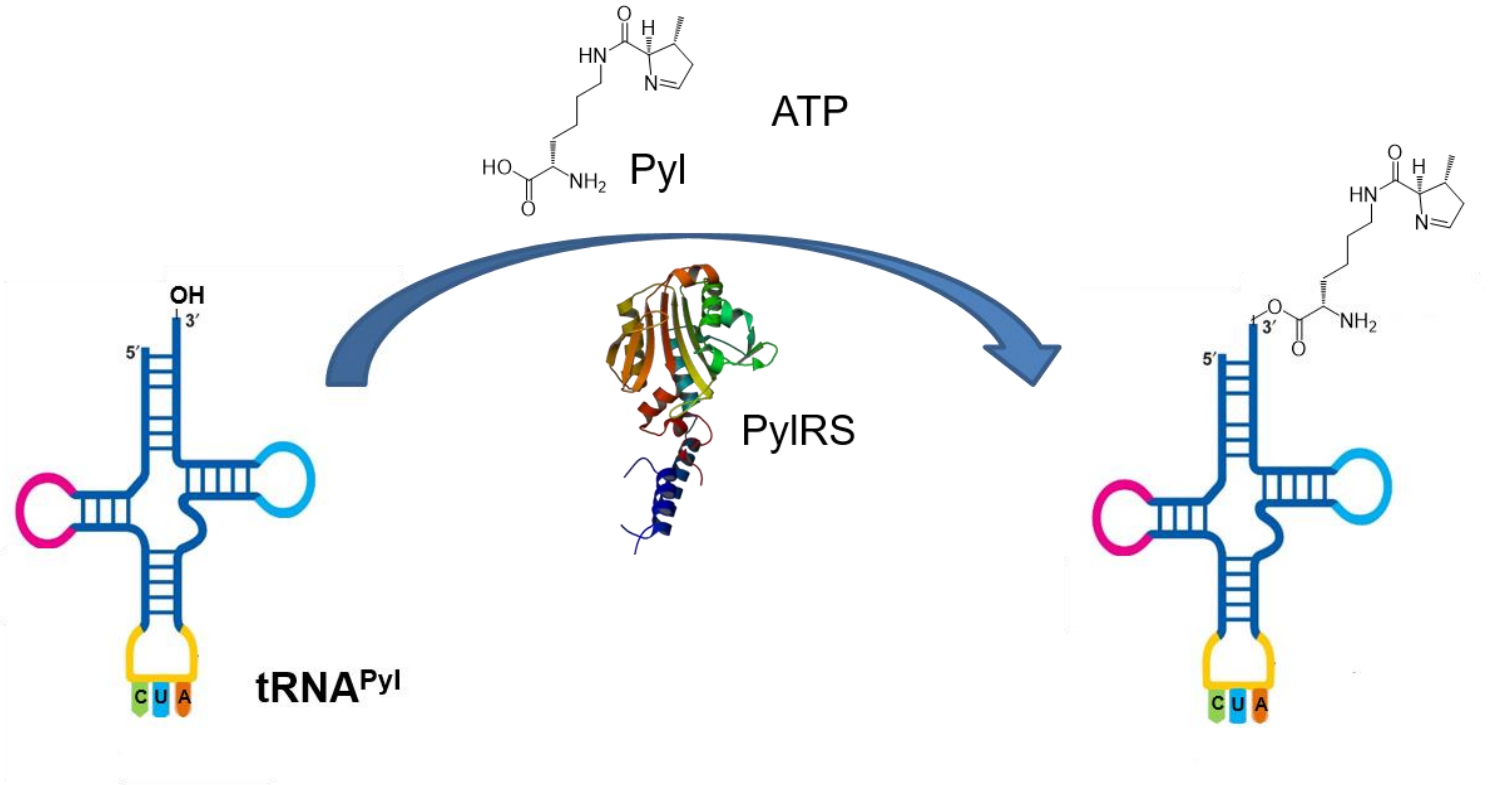
Sherry K. Blight[1]*, Ross C. Larue[1]*, Anirban Mahapatra[1]*, David G. Longstaff[1], Edward Chang[1], Gang Zhao[2]†, Patrick T. Kang[4], Kari B. Green-Church[5], Michael K. Chan[2,3,4] & Joseph A. Krzycki[1,4]

Nature **2004**, *431*, 333.

An aminoacyl-tRNA synthetase that specifically activates pyrrolysine

Carla Polycarpo*†, Alexandre Ambrogelly*†, Amélie Bérubé‡, SusAnn M. Winbush‡, James A. McCloskey§¶, Pamela F. Crain§, John L. Wood‡, and Dieter Söll*‡‖

PNAS **2004**, *101*, 12450.



◆ How Pyl is encoded

Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vitro* and *in vivo*

Sherry K. Blight[1*], Ross C. Larue[1*], Anirban Mahapatra[1*], David G. Longstaff[1], Edward Chang[1], Gang Zhao[2†], Patrick T. Kang[4], Kari B. Green-Church[5], Michael K. Chan[2,3,4] & Joseph A. Krzycki[1,4]

Nature **2004**, *431*, 333.

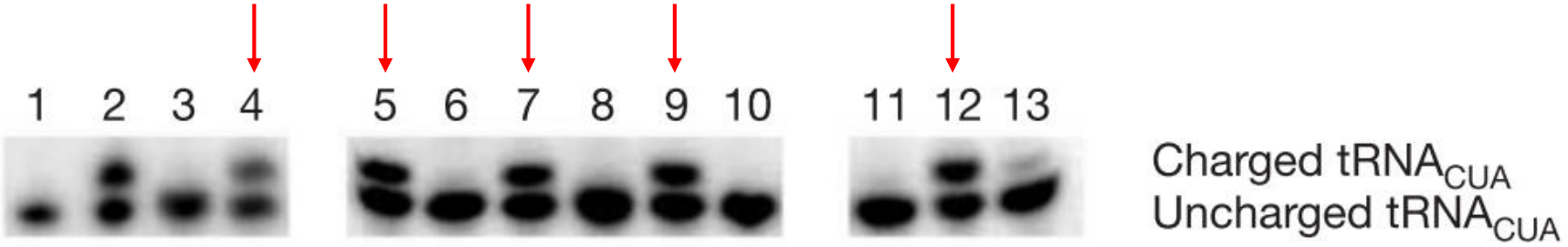
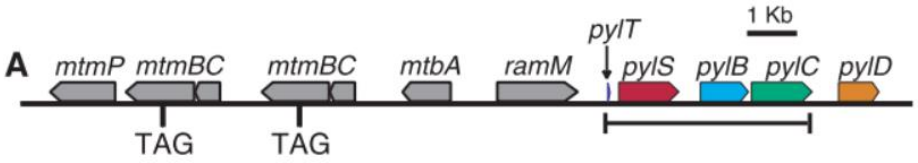


Figure 1. Aminoacylation of tRNACUA in cellular tRNA pools monitored by acid-urea gel electrophoresis and northern blotting to detect $tRNA_{CUA}$. **Lane 1** contains the alkali- deacylated cellular tRNA pool with only uncharged tRNACUA, **lane 2** contains cellular tRNA as isolated and shows charged (upper band) and uncharged (lower band) $tRNA_{CUA}$. Aminoacylation of $tRNA_{CUA}$ in the deacylated cellular tRNA pool by PylS was assayed as described in Methods except that the amino-acid substrate (50 uM) was varied. Each lane represents reactions with the following: **3**, no amino acid; **4**, pyrrolysine; **5**, pyrrolysine; **6**, lysine; **7**, pyrrolysine plus lysine; **8**, a mixture of the 20 canonical amino acids; **9**, pyrrolysine plus a mixture of the 20 canonical amino acids; **10**, pyrrolysine but lacking PylS-His6; **11**, no amino acid; **12**, pyrrolysine; **13**, the low-molecular-mass fraction from M. acetivorans cell extract. Sample groups 1–4 and 5–13 are from two different experiments. Samples 5–13 are from different parts of the same gel.

◆ Biosynthesis of Pyl



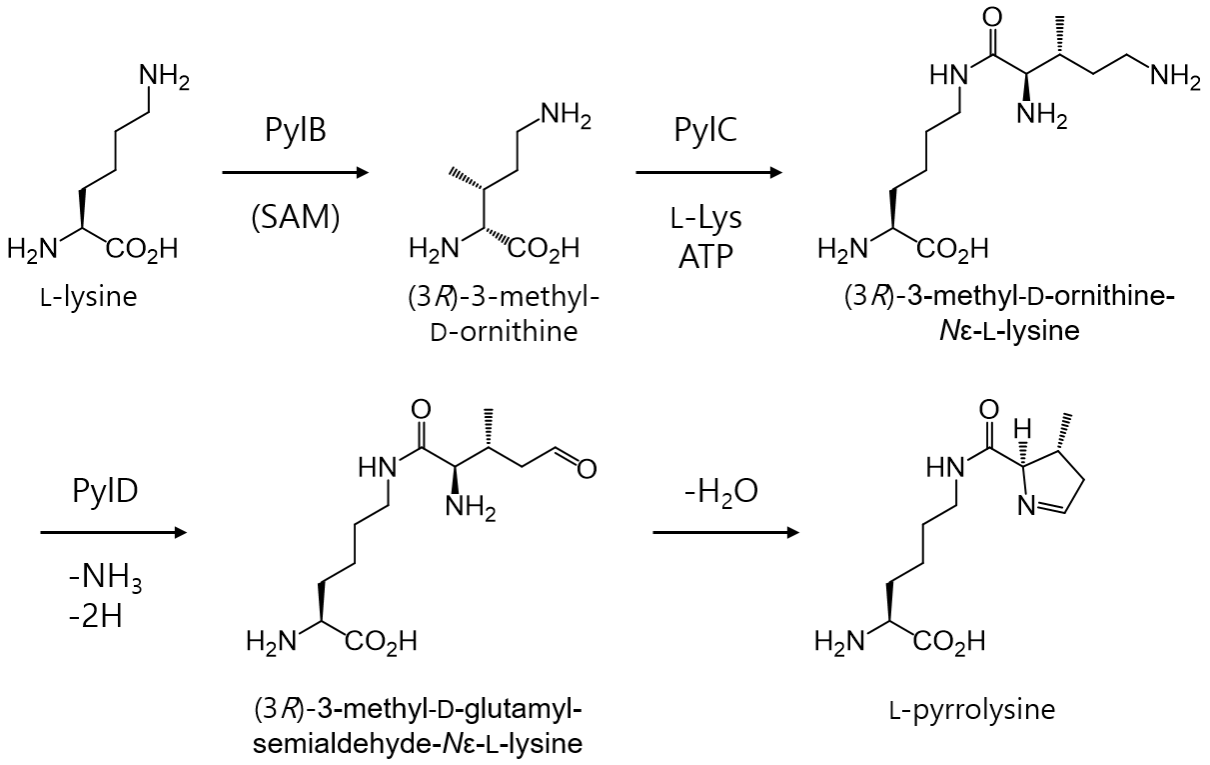
LETTER

doi:10.1038/nature09918

The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine

Marsha A. Gaston[1], Liwen Zhang[2], Kari B. Green-Church[2] & Joseph A. Krzycki[1,3]

Nature **2011**, *471*, 647.



◆ Methods for UAAs Incorporation

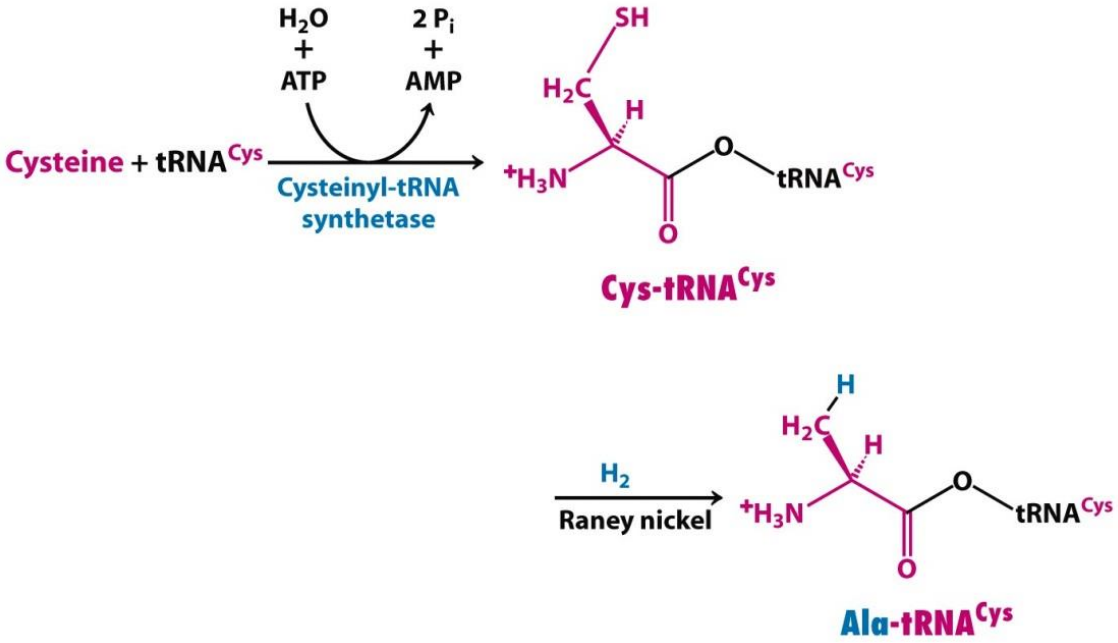
- Chemical modification of proteins
- Solid phase peptide synthesis
- Expressed Protein Ligation (EPL)
  - Ligation between synthetic peptides and native truncated proteins
- In vitro protein biosynthesis
  - Chemical acylation of UAAs to tRNA
- In vivo protein mutagenesis
  - Amino acid auxotrophs, synthetase specificity relaxation

◆ *IN VITRO* PROTEIN BIOSYNTHESIS

- "Adaptor hypothesis"
  - Anticodon–codon recognition between mRNA and tRNA is largely independent of the structure of the amino acid linked to the 3'-terminus of the acceptor stem of the tRNA
- The chemical modification of enzymatically amino-acylated tRNAs (aa-tRNAs) with synthetic reagents and probes
  - Treatment of Lys-$tRNA^{Lys}$ with N-acetoxysuccinimide affords (N-ε-acetyl-Lys)-$tRNA^{Lys}$, which can be used to insert N-ε-acetyl-Lys into proteins in a cell-free rabbit reticulocyte system
  - Low yield of aa-tRNA and multi-site incorporation

Codon-anticodon interactions

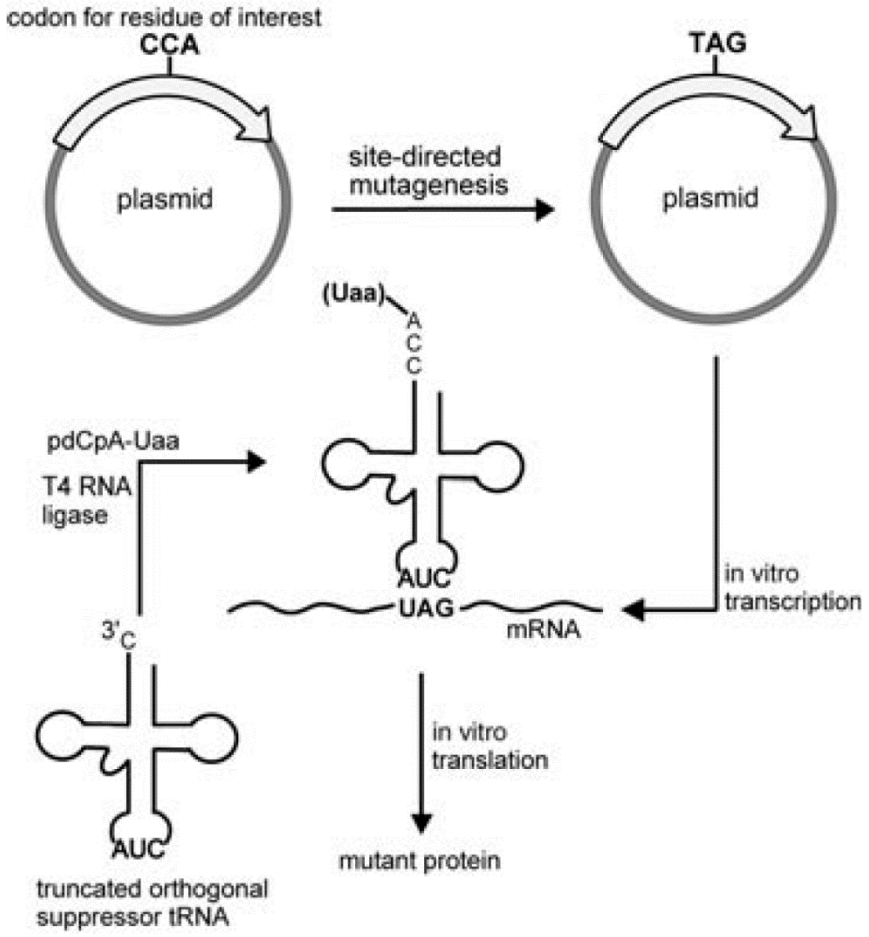
- Does the AA attached to the tRNA play any role?



- The AA in aminoacyl-tRNA does not play a role in selecting a codon

◆ *IN VITRO* PROTEIN BIOSYNTHESIS

- A general approach uses:
  - Enzymatic aminoacylation
  - Replacement of the codon for residue of interest with TAG(UAG) codon
  - Orthogonal $tRNA^{TAG}$
  - In vitro transcription and translation



◆ *IN VITRO* PROTEIN BIOSYNTHESIS

- Site-specifical incorporation of a large number of unnatural amino acids into various proteins
- Mutagenesis studies have provided important insights into protein structure and function
- Technically demanding, the low yields of mutant proteins, and hard to study proteins in cells
- The nature of the amino acid side chain is limited by the aminoacylation chemistry and the stability of the aa-tRNA linkage.

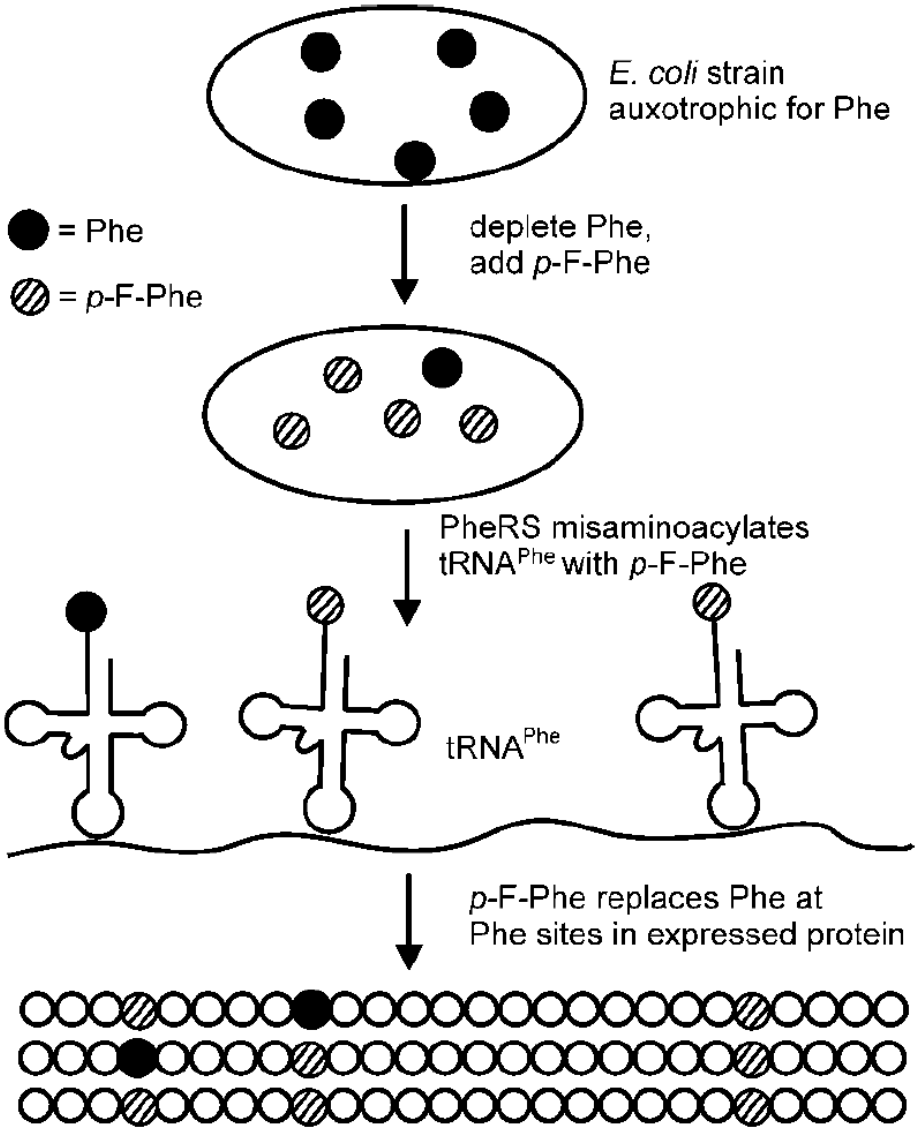
IN VIVO PROTEIN MUTAGENESIS

- The ability to incorporate unnatural amino acids directly into proteins in vivo has several advantages over in vitro methods
  - High yields of mutant proteins
  - Technical ease
  - The potential to study proteins in cells or multicellular organisms
- This can be accomplished by substitution of one of the common 20 amino acids with an unnatural amino acid
- The unnatural amino acid typically replaces the common amino acid throughout a protein (and the proteome), often in competition with a common amino acid, and results in heterogeneous proteins

◆ *IN VIVO* PROTEIN MUTAGENESIS

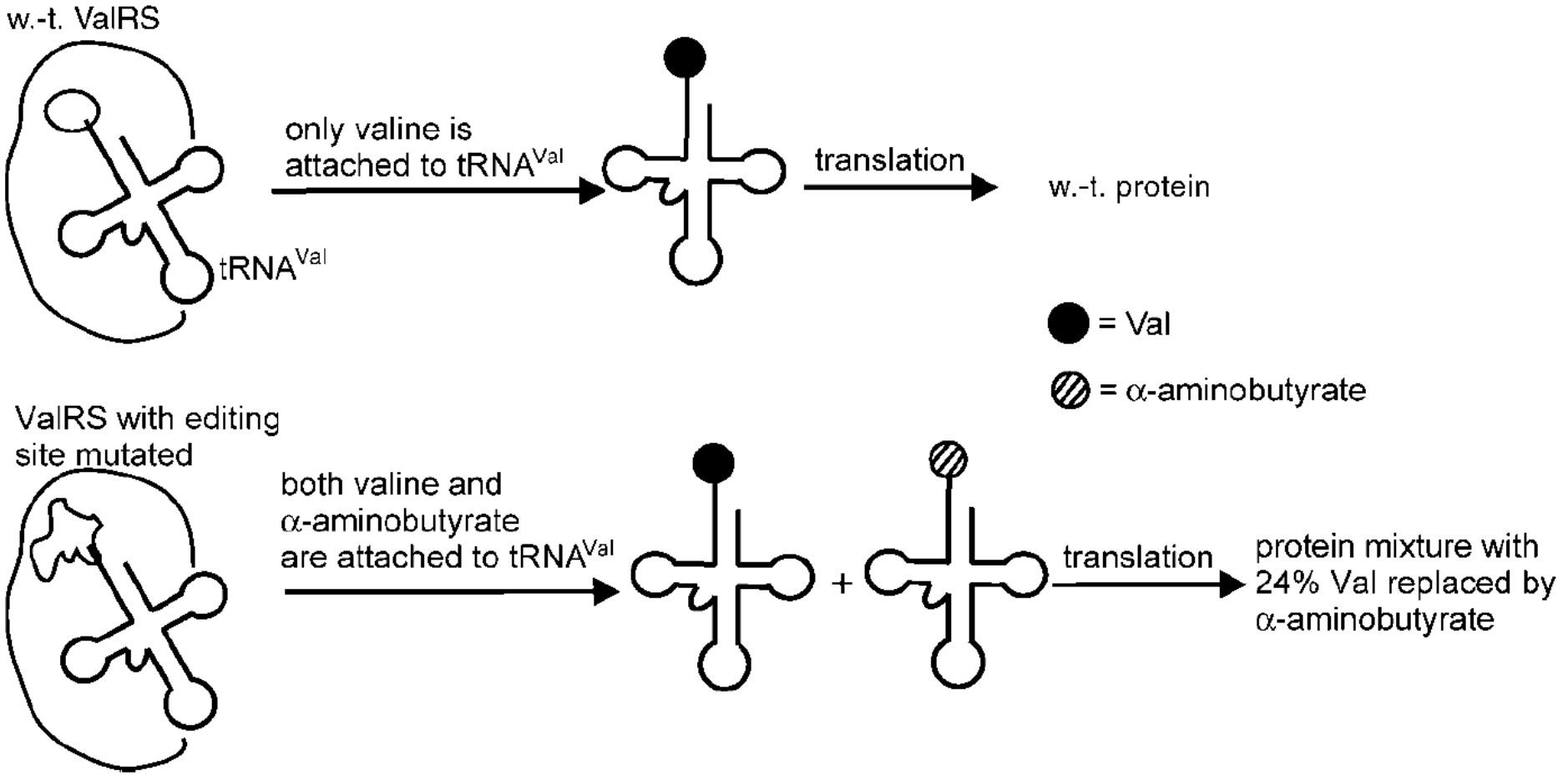
- Site-directed mutagenesis
  - The 20 natural AA can be introduced
- The use of auxotrophic bacterial strains to incorporate unnatural amino acids
  - Auxotrophy - inability of an organism to synthesize a particular organic compound required for its growth
- Relaxing the substrate specificity of aminoacyl-tRNA synthetases
- Attenuating the proofreading activity of aminoacyl-tRNA synthetases

◆ *IN VIVO* PROTEIN MUTAGENESIS



- Multiple substitution of phenylalanine by *p*-fluoro-phenylalanine (*p*-F-Phe) using a phenylalanine auxotrophic strain.

◆ *IN VIVO* PROTEIN MUTAGENESIS



Incorporation of α-aminobutyrate by a ValRS mutant with reduced proofreading activity. (w.-t.=wild-type)

IN VIVO PROTEIN MUTAGENESIS

- Though very useful, the use of auxotrophic strains for mutagenesis has a number of limitations
  - This method is not site-specific
  - The extent of incorporation of the unnatural amino acid is limited
  - Applicable only to close structural analogues of the common amino acids

◆ THE 1ST REPORT

Expanding the Genetic Code of *Escherichia coli*

Lei Wang,[1] Ansgar Brock,[2] Brad Herberich,[1] Peter G. Schultz[1,2]*

A unique transfer RNA (tRNA)/aminoacyl-tRNA synthetase pair has been generated that expands the number of genetically encoded amino acids in *Escherichia coli*. When introduced into *E. coli*, this pair leads to the in vivo incorporation of the synthetic amino acid *O*-methyl-L-tyrosine into protein in response to an amber nonsense codon. The fidelity of translation is greater than 99%, as determined by analysis of dihydrofolate reductase containing the unnatural amino acid. This approach should provide a general method for increasing the genetic repertoire of living cells to include a variety of amino acids with novel structural, chemical, and physical properties not found in the common 20 amino acids.

Schultz *et al.*, *Science* 2001.

◆ NEWS FOCUS IN *SCIENCE*

NEWS FOCUS

What would life look like if DNA contained more than four nucleotide bases and proteins more than 20 amino acids? Peter Schultz aims to find out

Creation's Seventh Day

In the casino or in the lab, Peter Schultz loves to take risks. "If I gamble, I usually gamble at high-stakes, high-payoff games," Schultz says. "Science is interesting when it's played at the same level, for the highest stakes with very high risk." For Schultz, a chemist at the Scripps Research Institute and the director of the newly created Genomics Institute of the Novartis Research Foundation (GNF), both in La Jolla, California, that betting system has paid handsomely.

While at his previous home at the University of California (UC), Berkeley, Schultz helped pioneer a fleet of high-speed chemistry techniques to generate molecules by the millions and select the ones that work best as possible catalysts, drug molecules, and even high-temperature superconductors. During the 1990s, he parlayed that experience into a string of start-up companies. Last year he took another big gamble by giving up the comfort of his Berkeley career and financial backing by the Howard Hughes Medical Institute to launch GNF, an outfit Schultz pitches as the "Bell Labs of Biology," which aims to work out the function of the thousands of unknown genes being turned out by the world's genome projects (see sidebar).

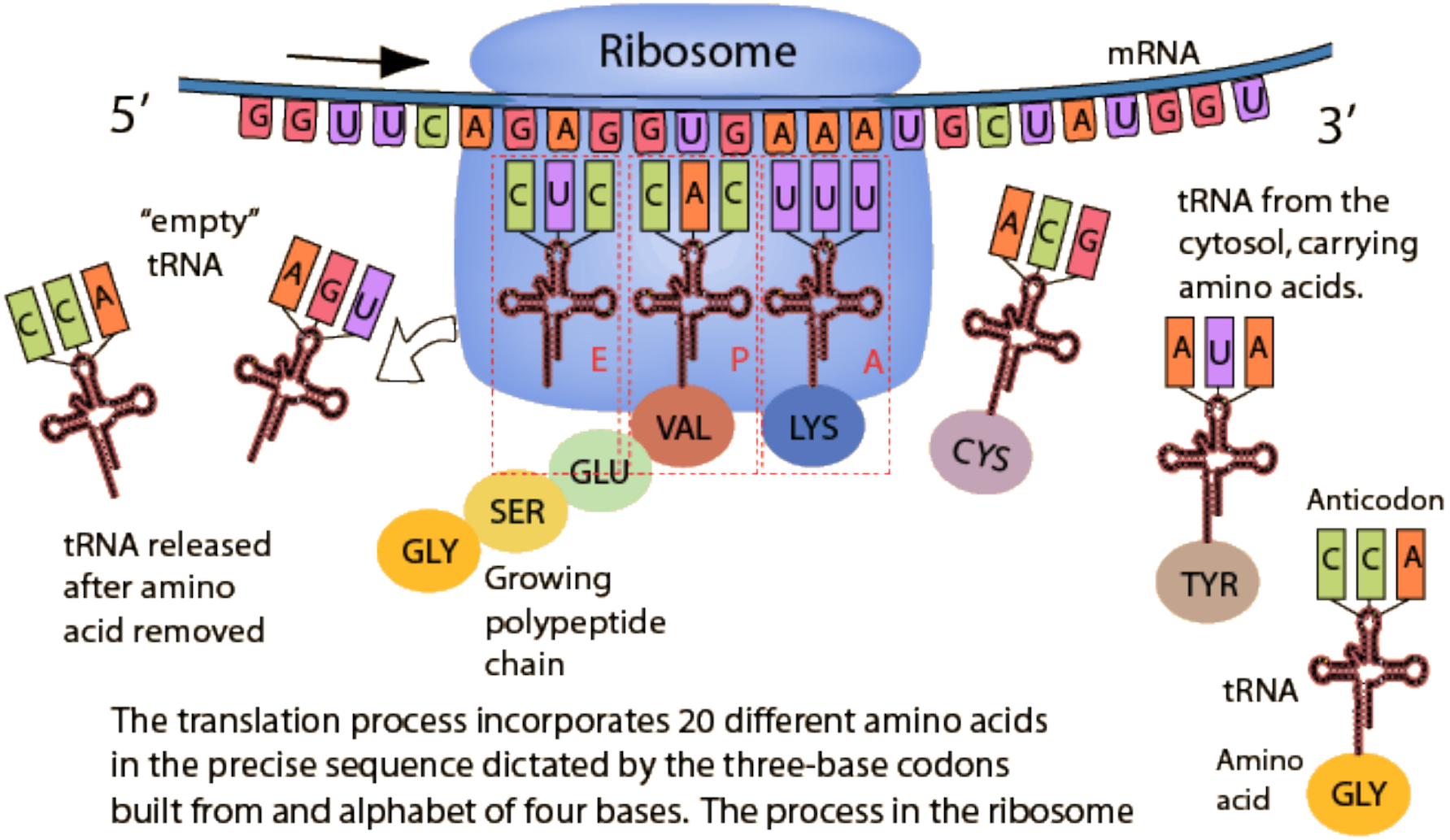
Still, his boldest undertaking—and the one that may ultimately have the highest impact—may lie in academic research: With colleagues at Scripps, Schultz is aiming to rewrite the basic chemistry of life. By reengineering DNA, RNA, and the proteins that interact with them, they hope to create synthetic organisms with a chemical makeup fundamentally different from all life that has existed on Earth for the last 3.8 billion years.

If they succeed, their biochemical reengineering could have a profound effect on everything from basic molecular biology to industrial chemistry. The result—they hope—will be proteins that incorporate amino acids other than the 20 commonly used by life to construct proteins. By adding these amino acids with completely new types of chemical behaviors, Schultz and his colleagues hope to design bacteria to make proteins that work as novel catalysts and drugs, or that ca
researchers decip
thrust of the wor
way genetic di
Steven Briggs, w
cultural Discover
zation to GNF.
work—and I'm s
a much bigger t
and other things t

It could also o
lution, allowing r
native paths life
its infancy and th
where in the ga
Steven Benner, a
from the Univers
allows researche
time whether an a
is truly viable. A
had worked a se
look like today?"

As scientists,

PROTEINS – BIOSYNTHESIS

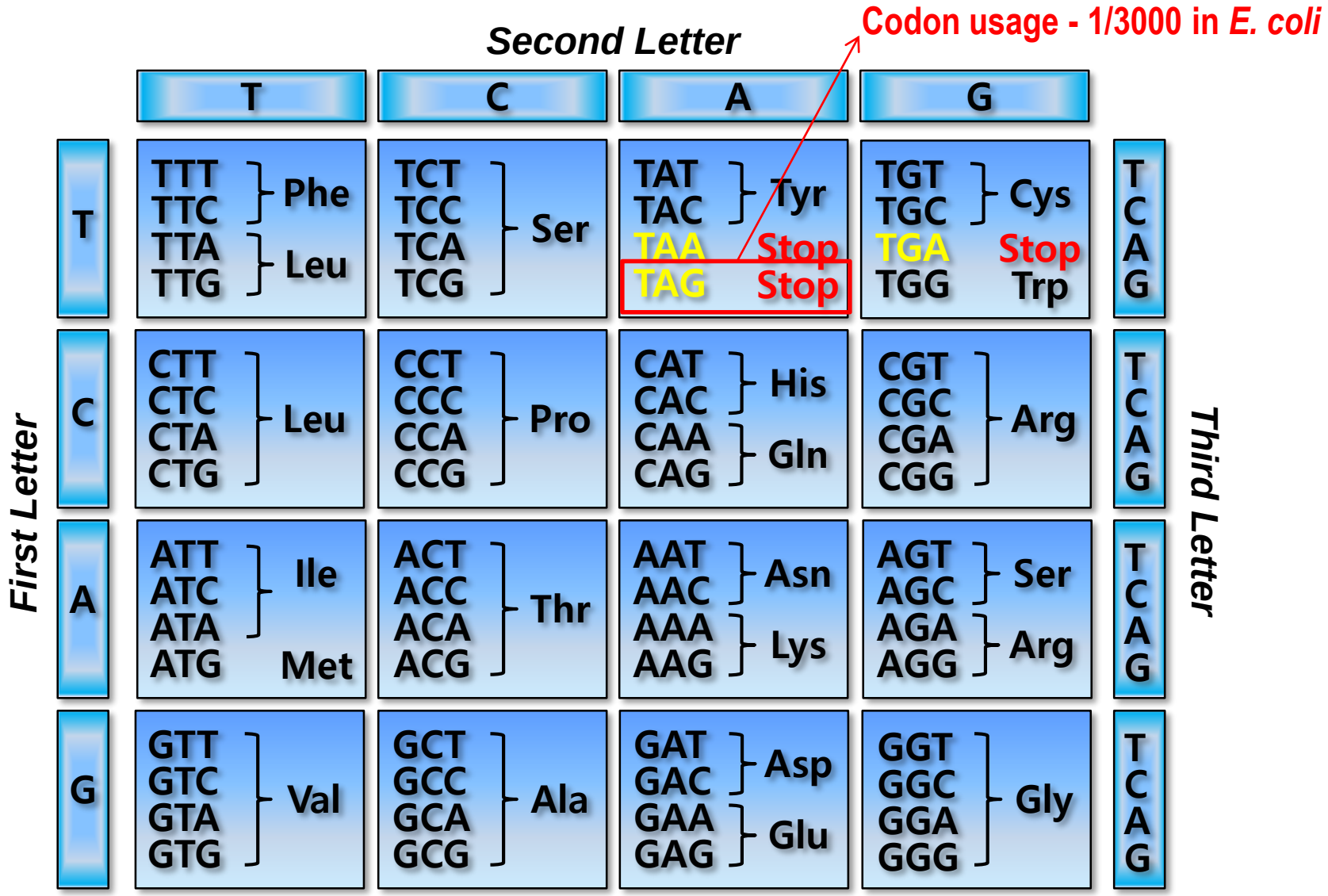


The translation process incorporates 20 different amino acids in the precise sequence dictated by the three-base codons built from and alphabet of four bases. The process in the ribosome builds the polypeptide chains tha will become proteins.

KEY COMPONENTS

- Requirements for the genetic incorporation
  - UAA
  - Unique codon – TAG (1/3000 in *E. coli*)
  - Orthogonal tRNA / synthetase pair
  - General method to evolve synthetases

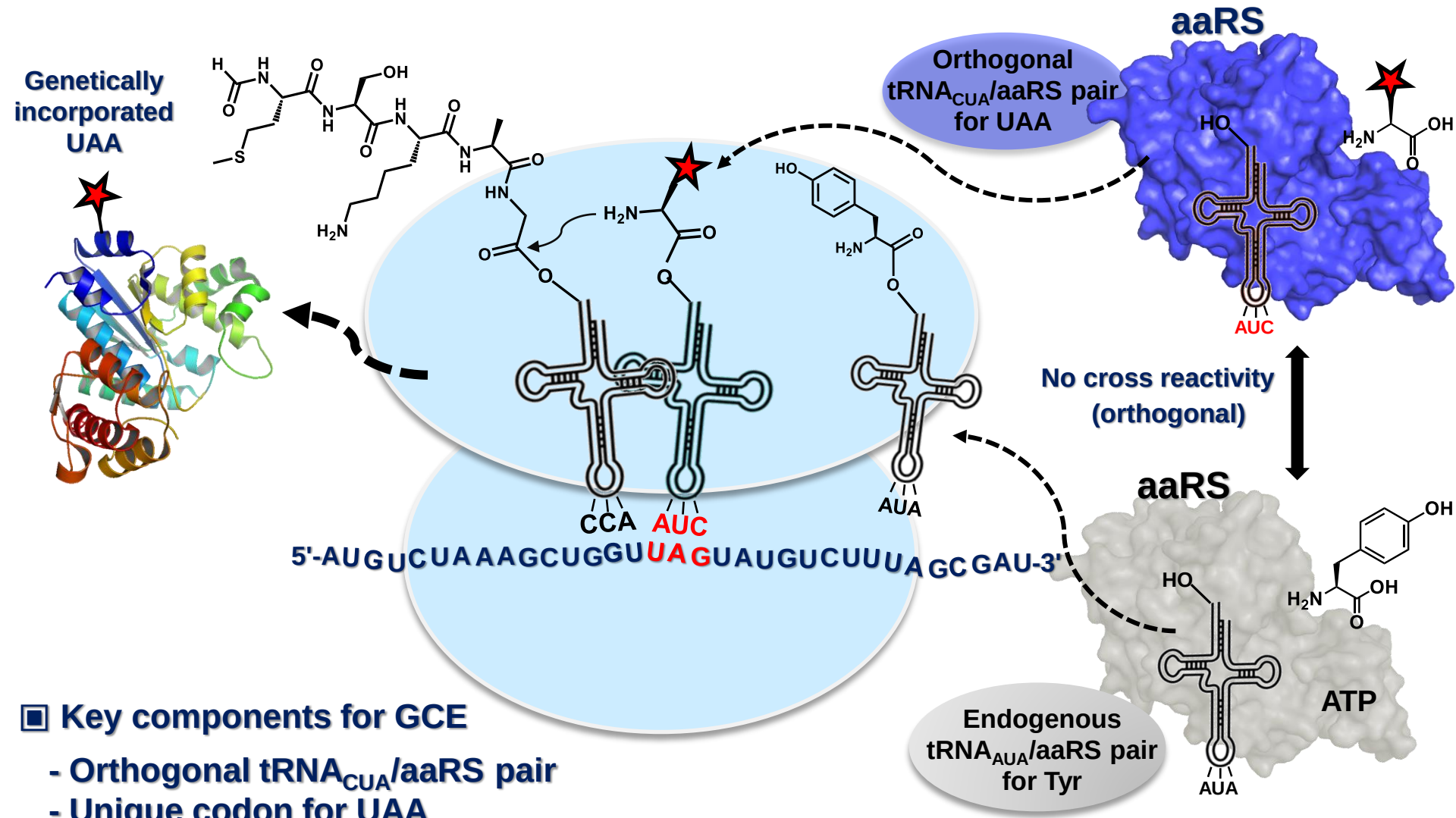
◆ UNIQUE CODON



GENETIC INCORPORATION OF UAAS

- Requirements for the genetic incorporation
  - UAA
  - Unique codon – TAG (1/3000 in *E. coli*)
  - Orthogonal tRNA / synthetase pair
  - General method to evolve synthetases

◆ Orthogonal tRNA / synthetase pair

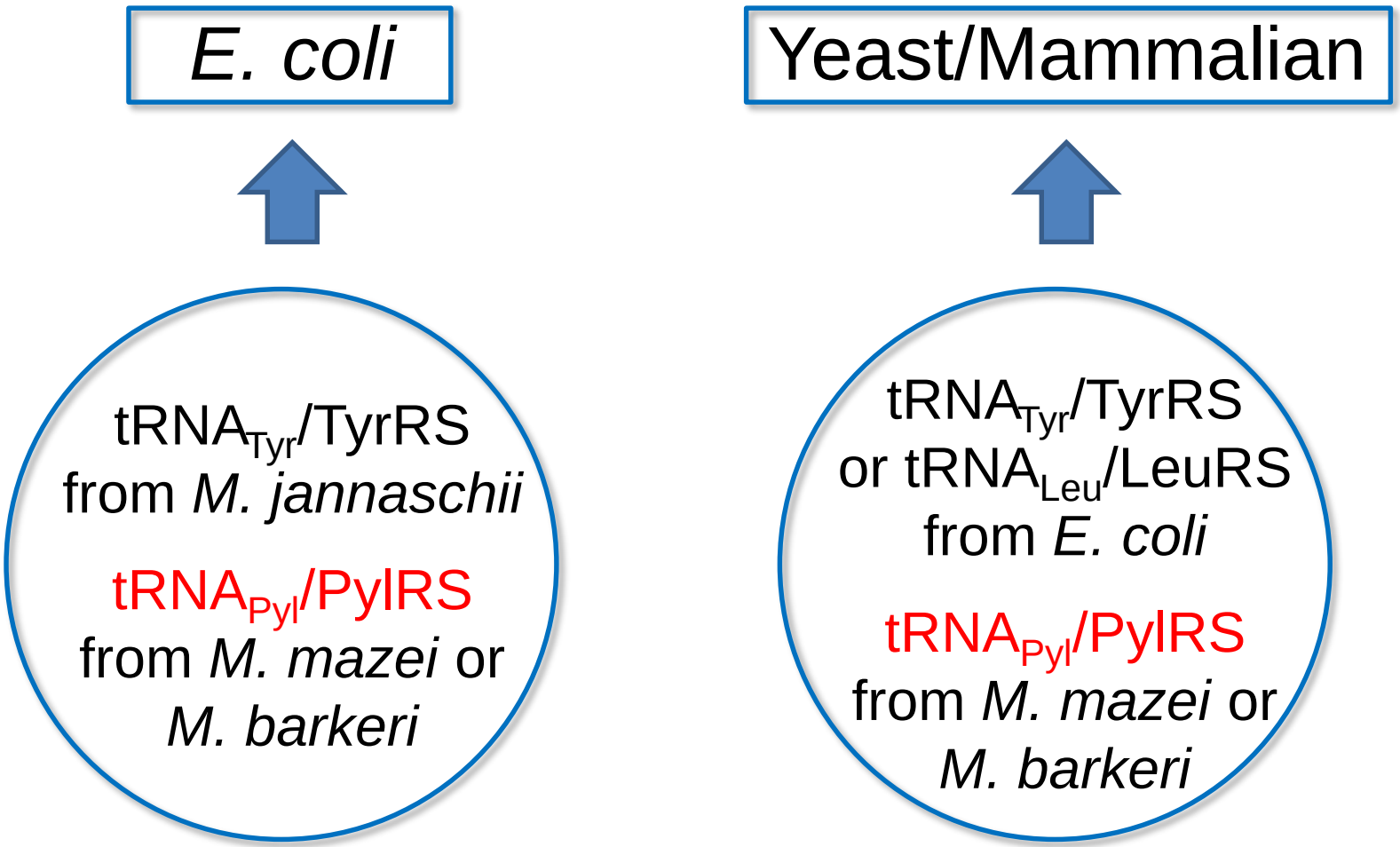


▣ Key components for GCE

- Orthogonal tRNA$_{CUA}$/aaRS pair
- Unique codon for UAA
- UAA (supplemented)

◆ Orthogonal tRNA / synthetase pair

- **Popular orthogonal tRNA / synthetase pairs**

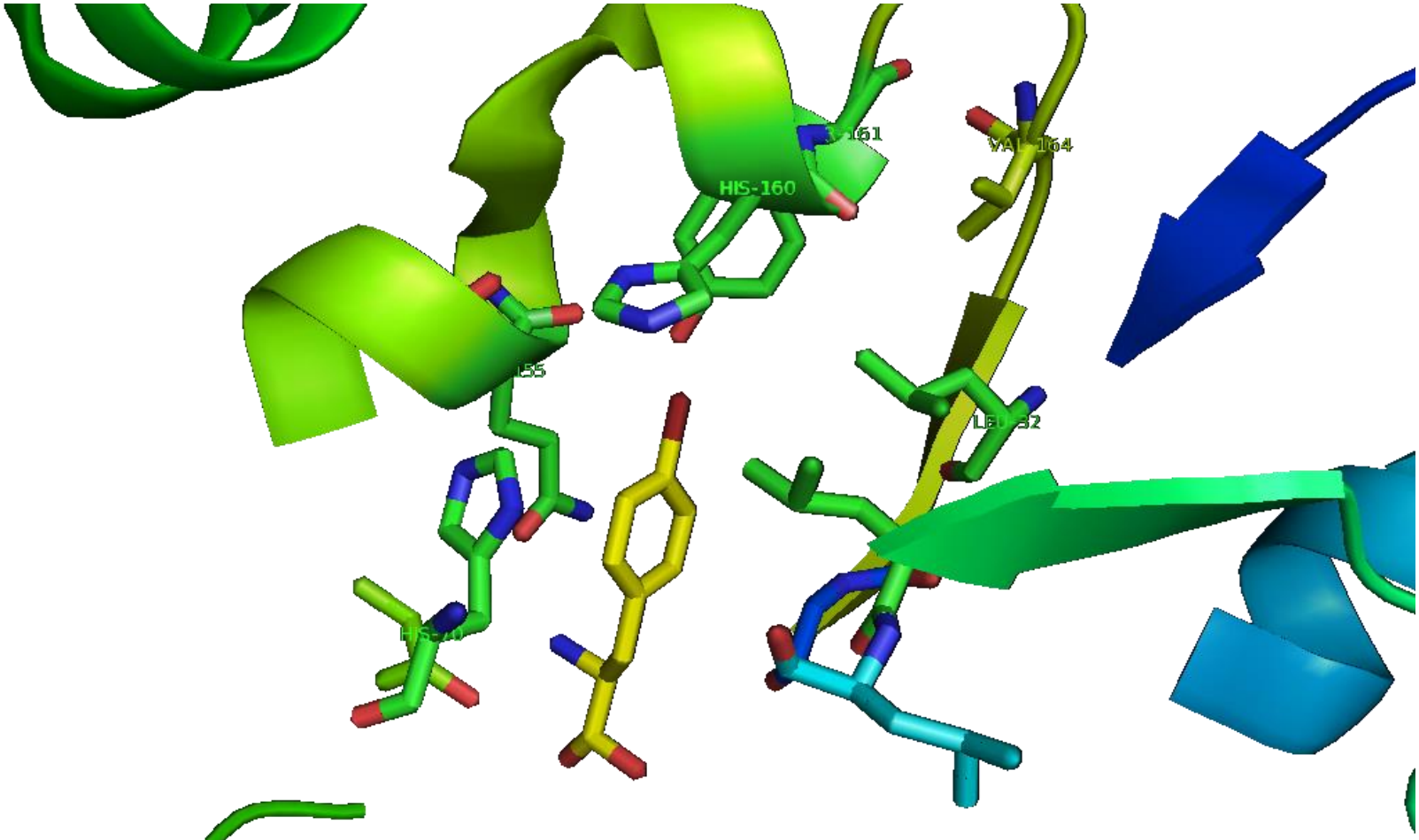


GENETIC INCORPORATION OF UAAS

- Requirements for the genetic incorporation
  - Unnatural amino acid
  - Unique codon – TAG (1/3000 in *E. coli*)
  - Orthogonal tRNA / synthetase pair
  - General method to evolve synthetases

ENGINEERING SYNTHETASES

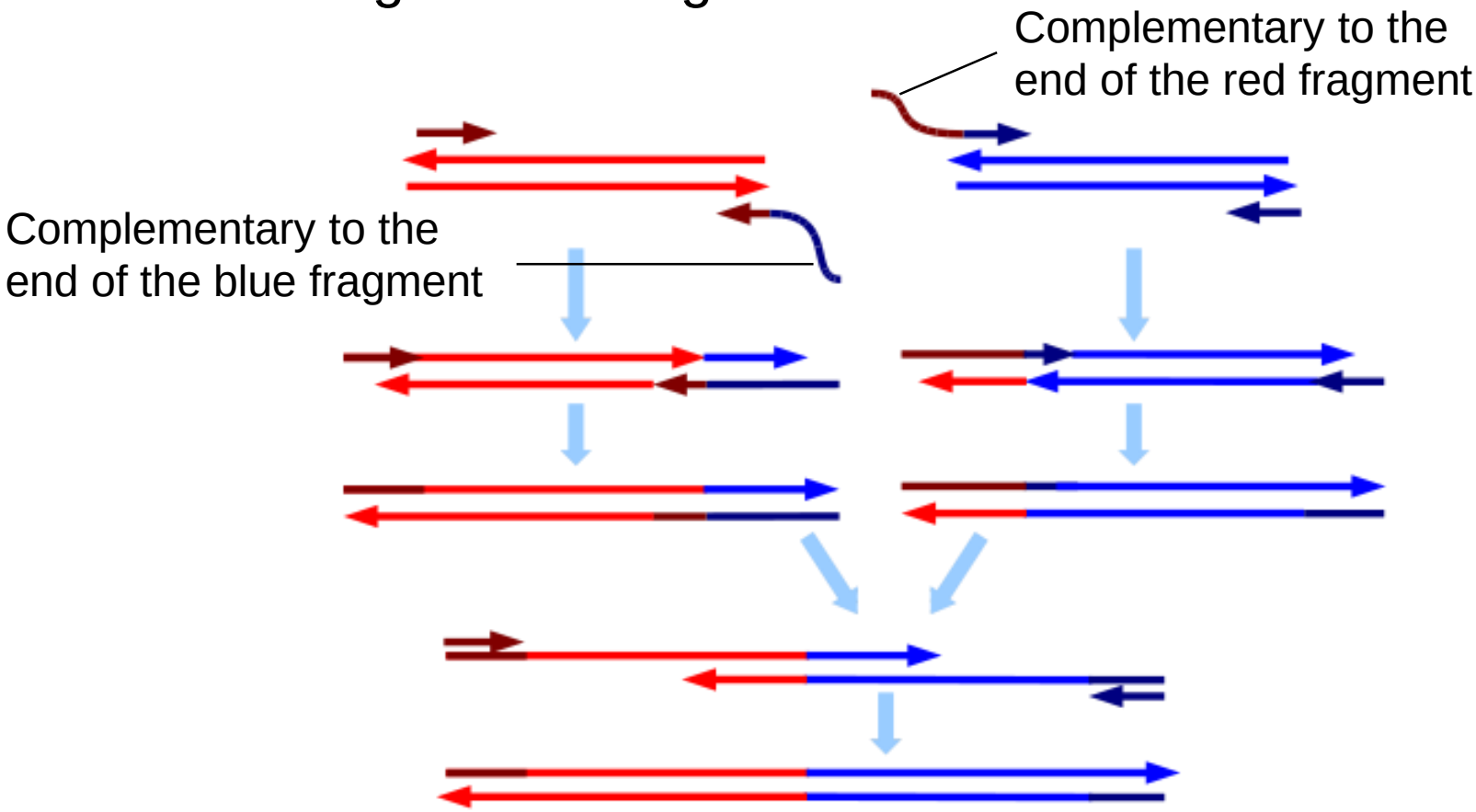
- **Engineering the active site of a synthetase**



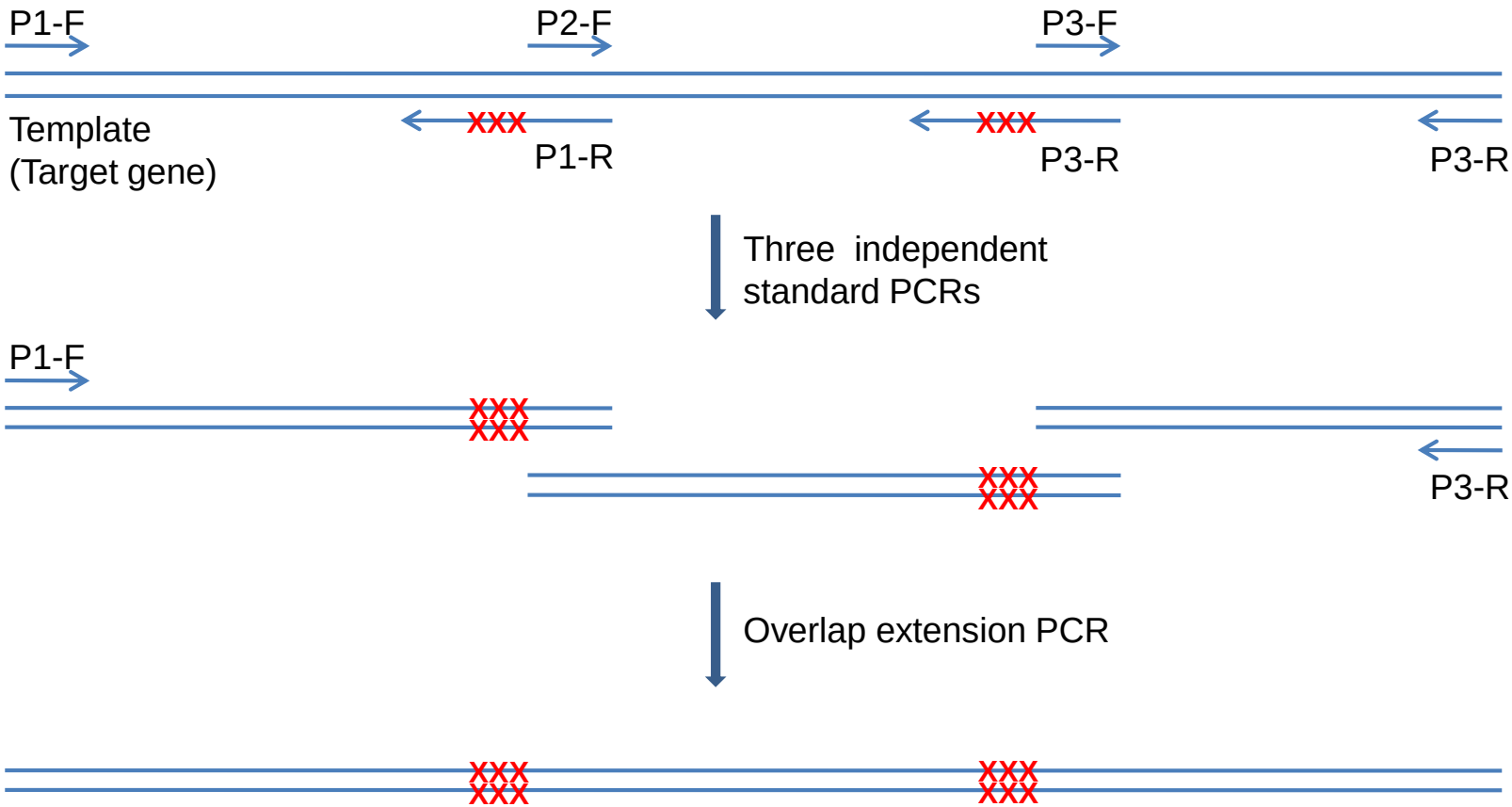
Amino Acid Binding Site of *Mj*TyrRS

Overlap Extension PCR

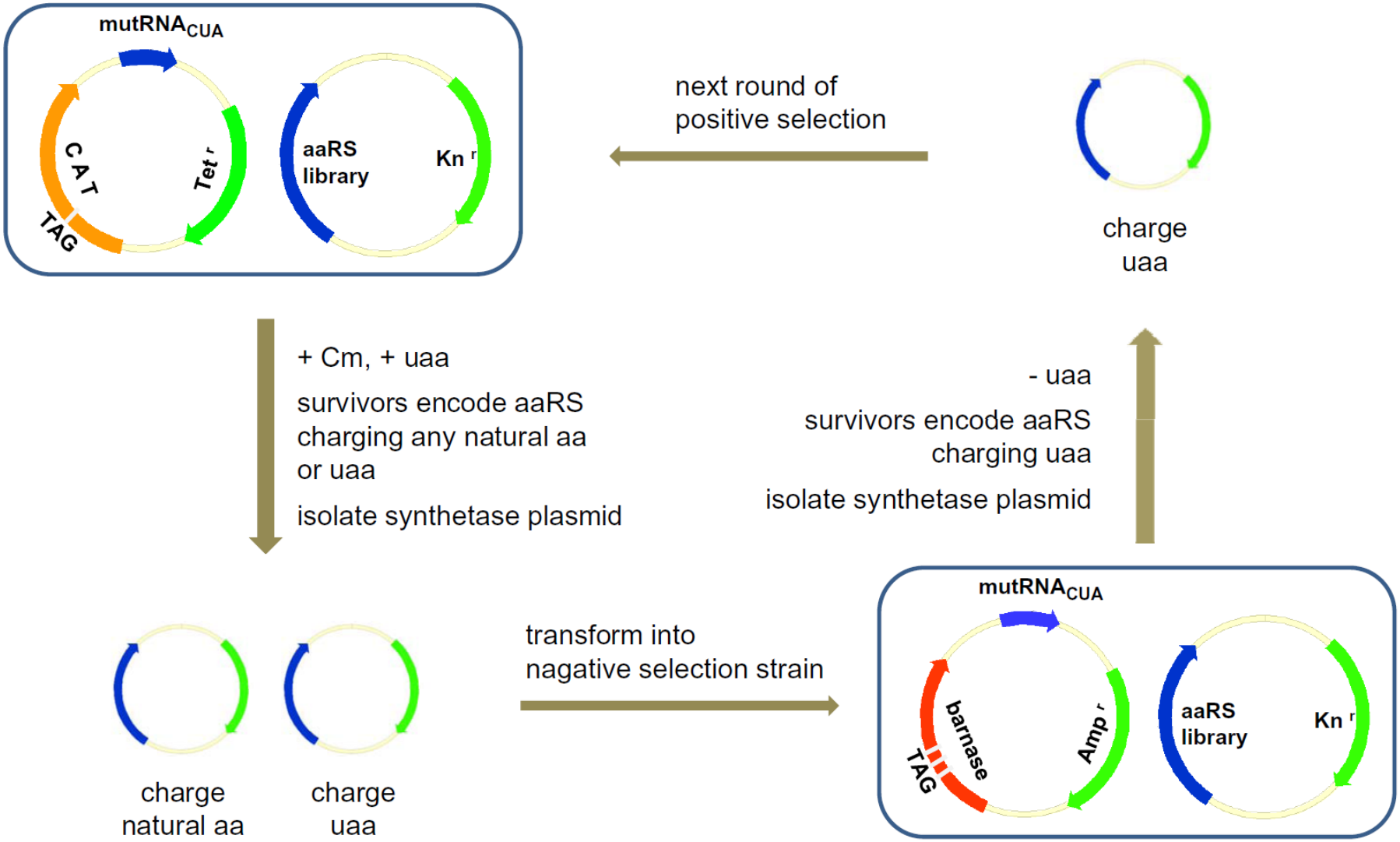
- It is used to insert specific mutations at specific points in a sequence or to combine smaller DNA fragments into a larger DNA fragment



◆ Overlap Extension PCR (optional)



Selection Scheme for *E. coli*



◆ THE 1ST REPORT

Expanding the Genetic Code of *Escherichia coli*

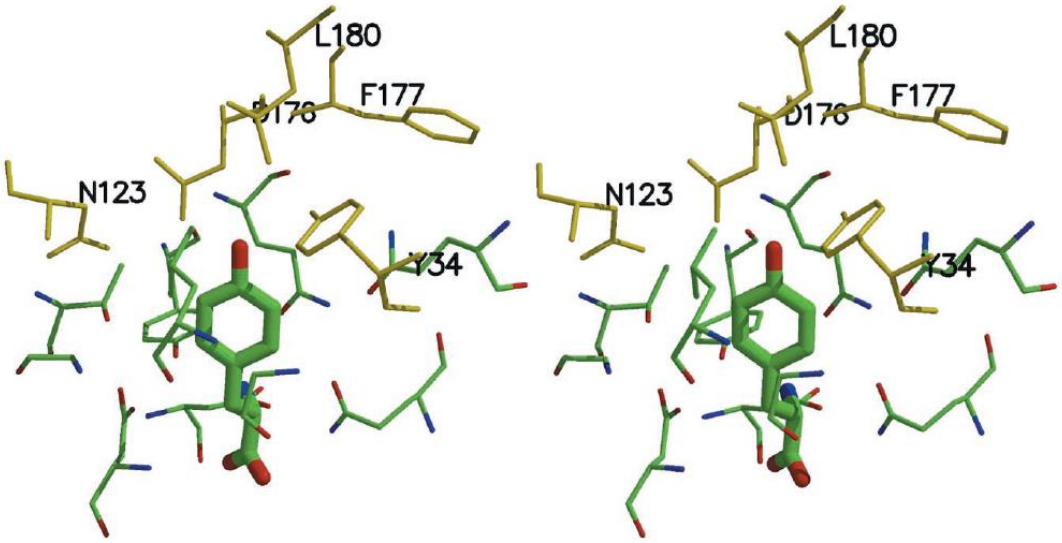


Fig. 1. Stereo view of the active site of TyrRS (*18*). Residues from *B. stearothermophilus* TyrRS are shown in the figure. Corresponding residues from *M. jannaschii* TyrRS are Tyr^{32} (Tyr^{34}), Glu^{107} (Asn^{123}), Asp^{158} (Asp^{176}), Ile^{159} (Phe^{177}), and Leu^{162} (Leu^{180}) with residues from *B. stearothermophilus* TyrRS in parenthesis; mutated residues are in yellow.

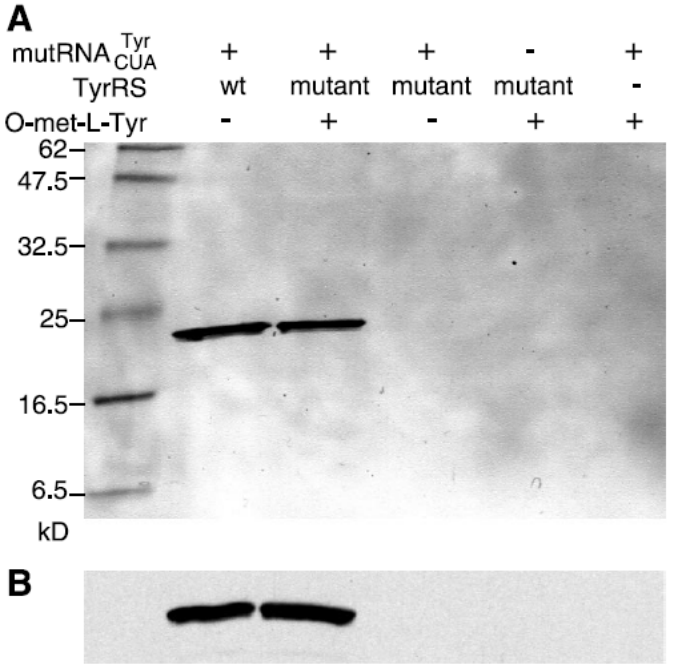


Fig. 2. Accumulation of *E. coli* DHFR protein under different conditions. (**A**) Silver-stained SDS-PAGE gel of purified DHFR. A six-histidine tag was added to the COOH terminus of *E. coli* DHFR, and protein was purified by immobilized metal affinity chromatography. Expression conditions are notated at the top of each lane. The left lane is a molecular weight marker. (**B**) Western blot of gel in (A). A penta-His antibody was used to detect the six-histidine tag at the COOH terminus of DHFR.

Science 2001.

◆ EXPANSION TO EUKARYOTES

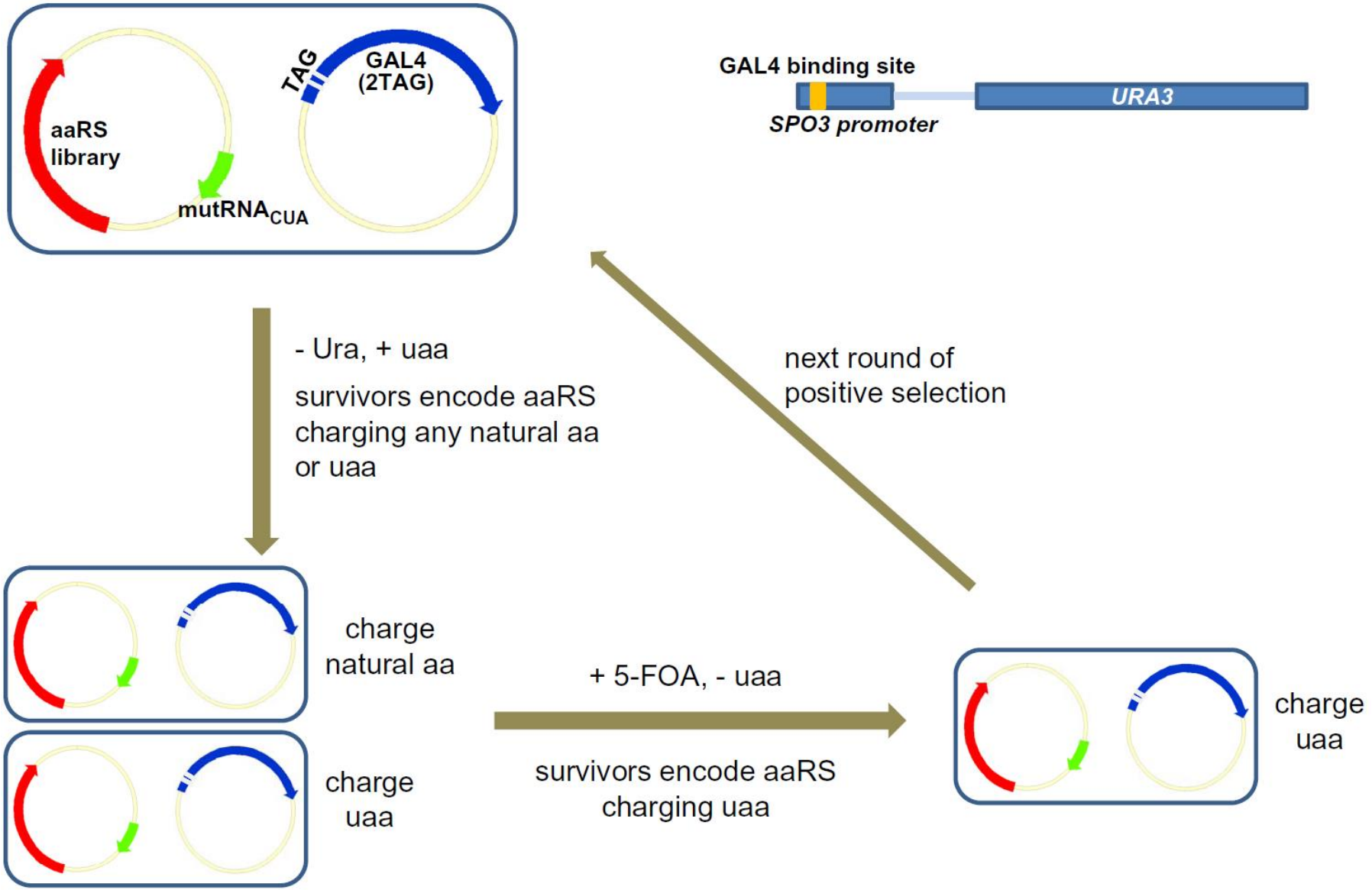
An Expanded Eukaryotic Genetic Code

Jason W. Chin,* T. Ashton Cropp, J. Christopher Anderson, Mridul Mukherji, Zhiwen Zhang, Peter G. Schultz†

We describe a general and rapid route for the addition of unnatural amino acids to the genetic code of *Saccharomyces cerevisiae*. Five amino acids have been incorporated into proteins efficiently and with high fidelity in response to the nonsense codon TAG. The side chains of these amino acids contain a keto group, which can be uniquely modified in vitro and in vivo with a wide range of chemical probes and reagents; a heavy atom–containing amino acid for structural studies; and photocrosslinkers for cellular studies of protein interactions. This methodology not only removes the constraints imposed by the genetic code on our ability to manipulate protein structure and function in yeast, it provides a gateway to the systematic expansion of the genetic codes of multicellular eukaryotes.

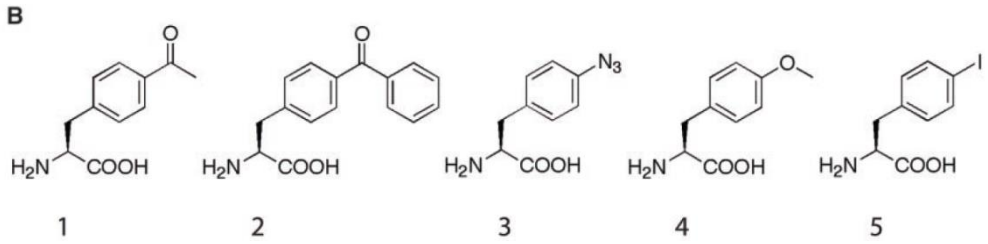
15 AUGUST 2003 VOL 301 SCIENCE

◆ Selection Scheme for *Yeast*



◆ EXPANSION TO EUKARYOTES

An Expanded Eukaryotic Genetic Code



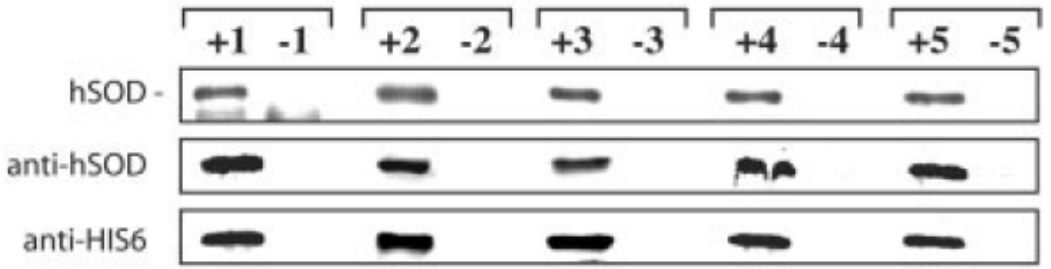
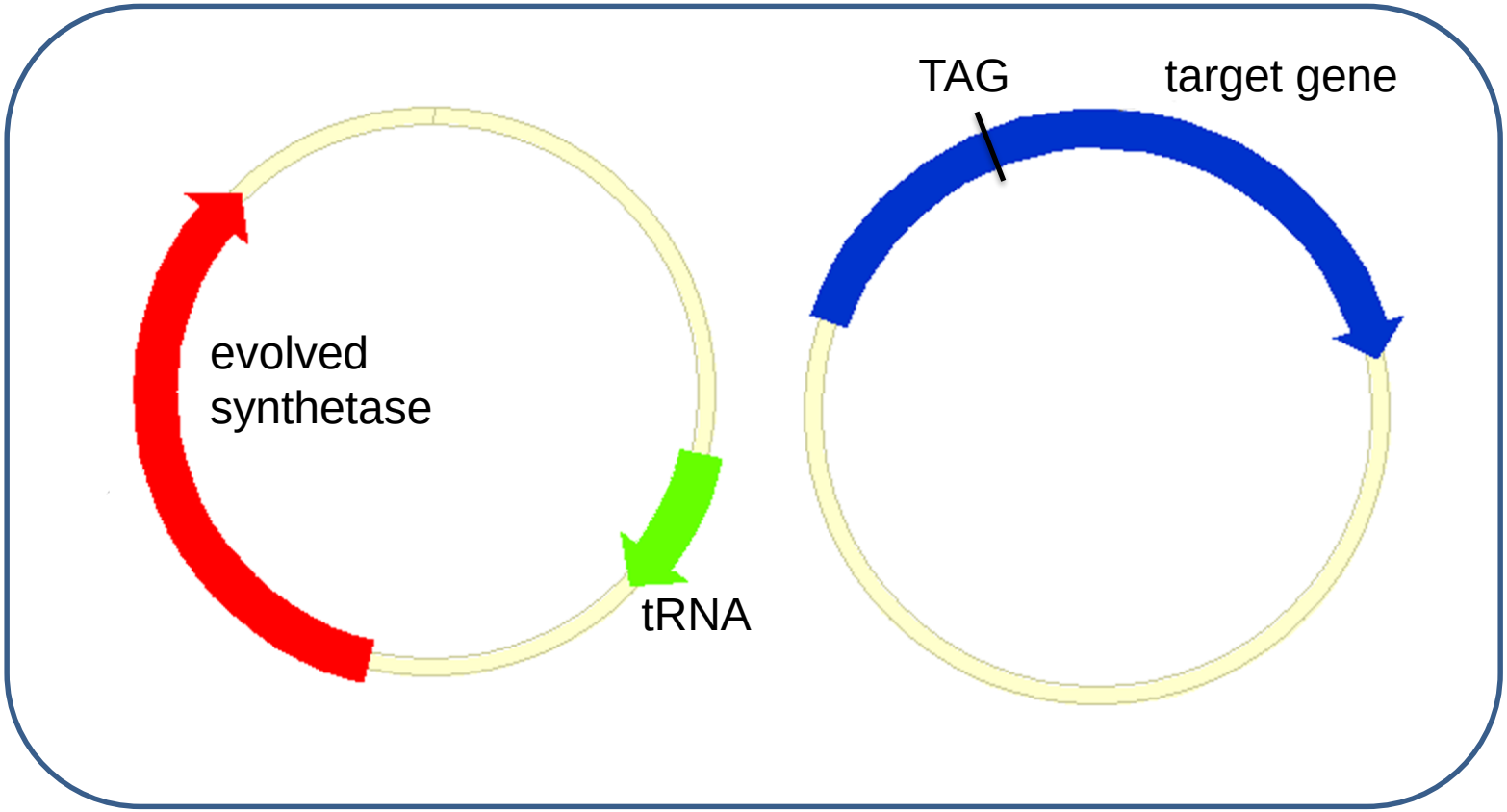


Fig. 3. Protein expression of hSOD (33TAG)HIS in *S. cerevisiae* genetically encoding unnatural amino acids. (**Top**) SDS–polyacrylamide gel electrophoresis of hSOD purified from yeast in the presence (+) and absence (–) of the unnatural amino acid indicated (see Fig. 1B) stained with Coomassie. Cells contain the mutant synthetase-tRNA pair selected for the amino acid indicated. (**Center**) Western blot probed with an antibody against hSOD. (**Bottom**) Western blot probed with an antibody against the C-terminal His6 tag.

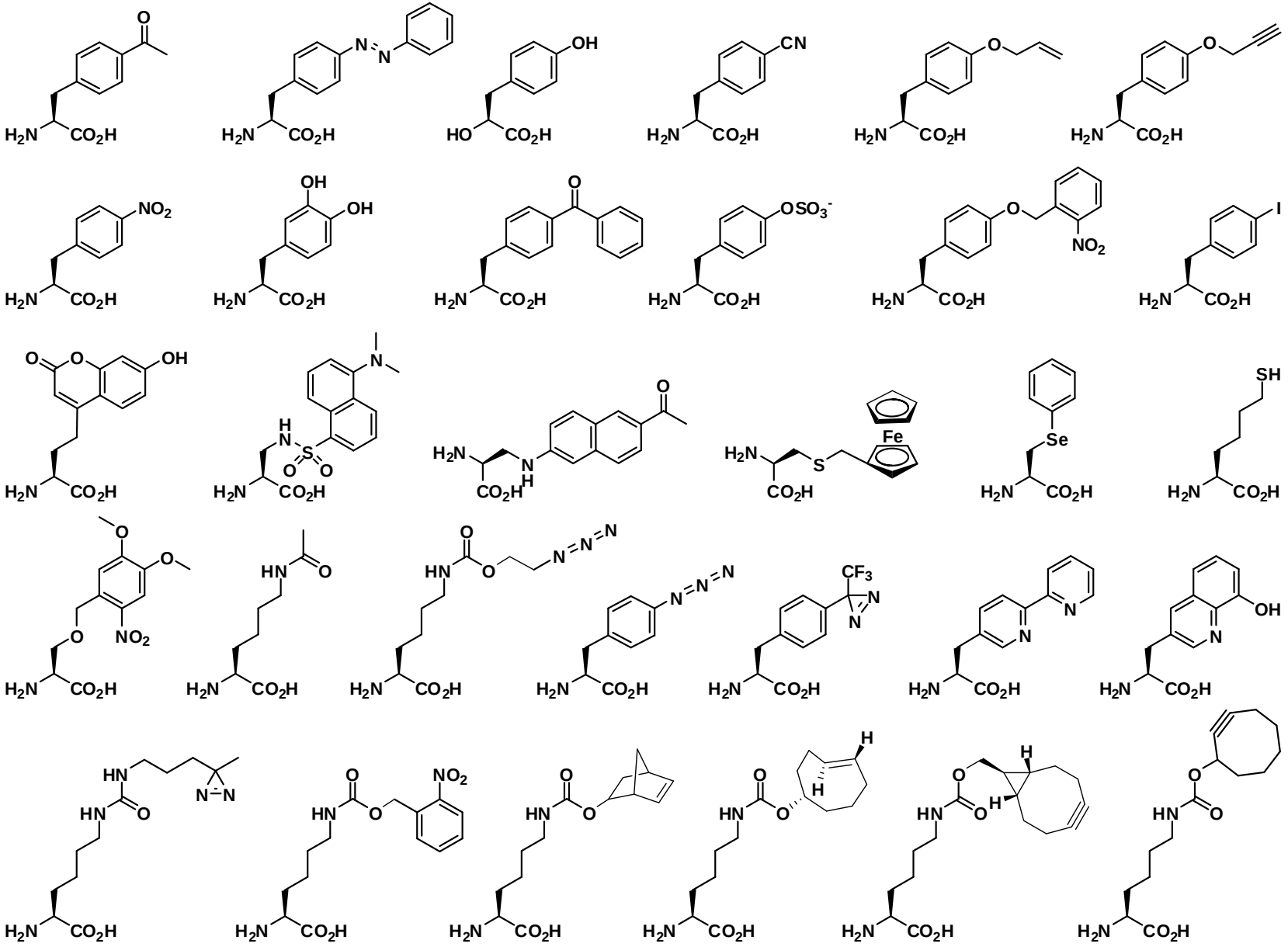
◆ HOW TO EXPRESS PROTEINS CONTAINING UAAS

UAA



UAA

GENETICALLY ENCODED UAAS



◆ PylT and PylRS as an universal system for GCE

- GCE technology requires orthogonality of an aatRNA/aaRS pair for operation in a host cell
- The orthogonality is usually obtained by applying an exogenous pair
- The $tRNA_{Pyl}$/PylRS pair is orthogonal in both prokaryotes and eukaryotes

E. coli

$tRNA_{Tyr}$/TyrRS from *M. jannaschii*

$tRNA_{Pyl}$/PylRS from *M. mazei* or *M. barkeri*

Mammalian cells

$tRNA_{Tyr}$/TyrRS or $tRNA_{Leu}$/LeuRS from *E. coli*

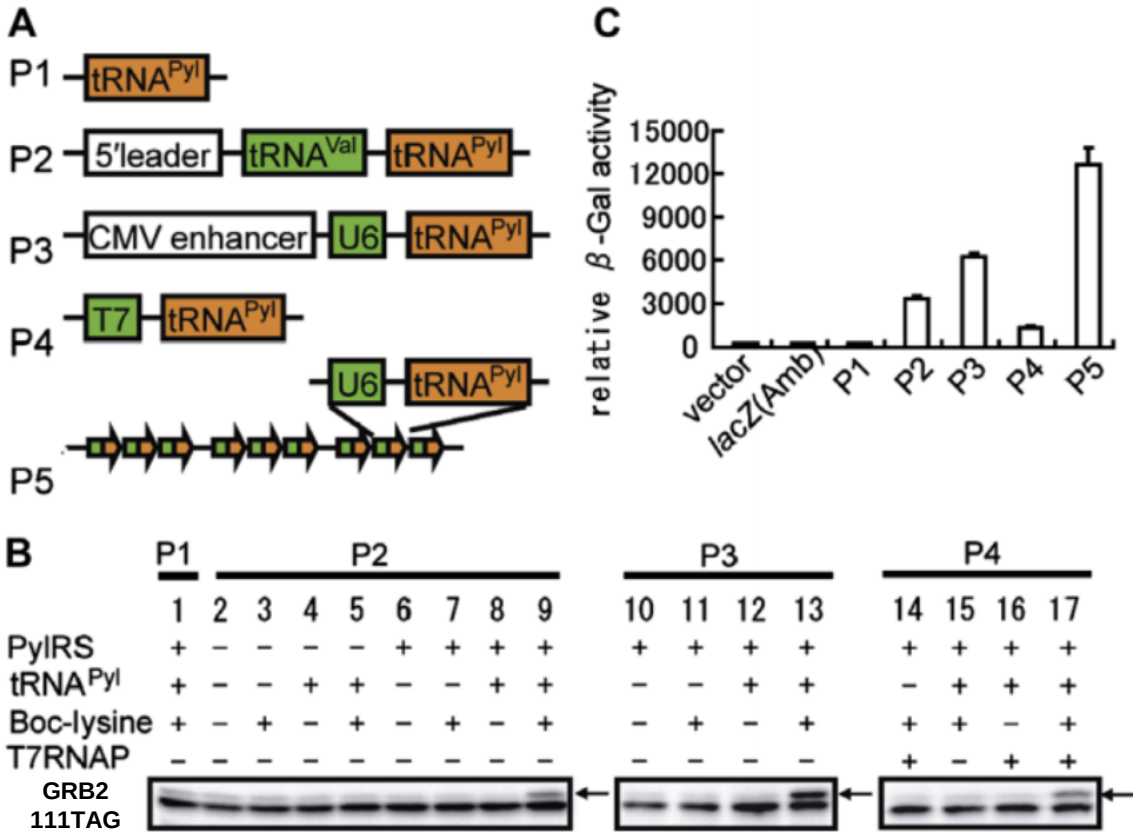
$tRNA_{Pyl}$/PylRS from *M. mazei* or *M. barkeri*

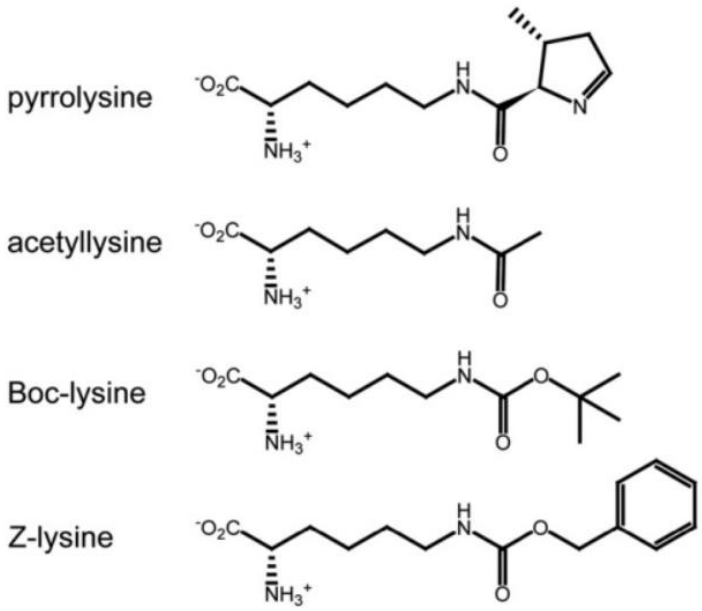
◆ PylT and PylRS as an universal system for GCE

Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases

Takahito Mukai [a,b,1], Takatsugu Kobayashi [a,1], Nobumasa Hino [a], Tatsuo Yanagisawa [a], Kensaku Sakamoto [a], Shigeyuki Yokoyama [a,b,*]

Biochem. Biophys. Res. Commun. **2008**, *371*, 818.





PylT and PylRS as an universal system for GCE