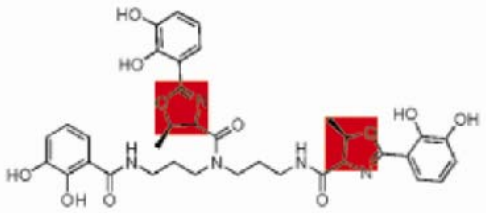
Non-Ribosomal Peptide Synthesis

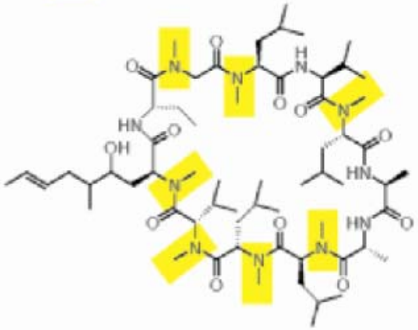
In contrast to proteins produced by ribosomal synthesis, many small peptide natural products contain not only the common 20 amino acids but also hundreds of different amino acids. These peptidic natural products are assembled by large enzymes, referred to as nonribosomal peptide synthetases (NRPS). These produce both cyclic and linear polypeptides which can contain non-proteinogenic amino acids. Some examples are:

heterocyclization



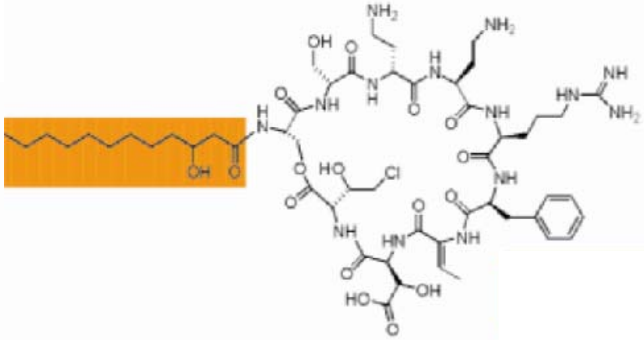
Vibriobactin

N-methylation



Cyclosporin

fatty acids

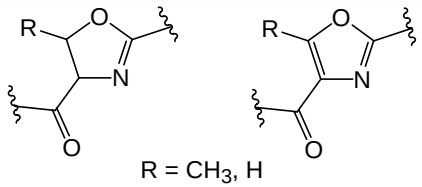


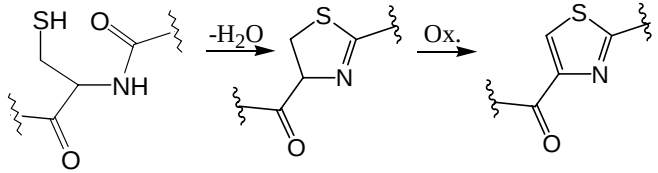
Syringomycin

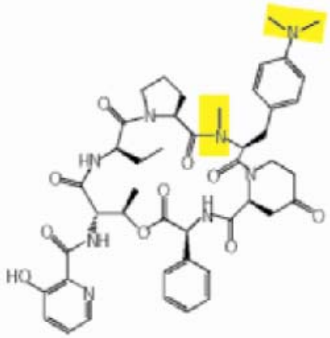
Carboxy acids and sugars

Bacillibactin

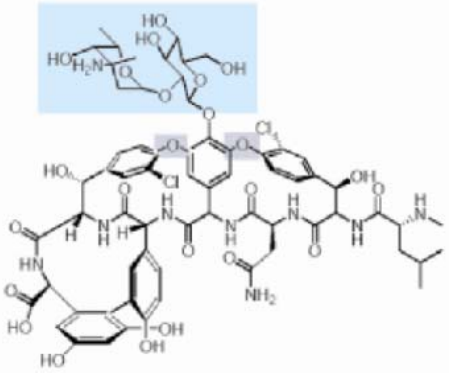
Typical examples:







Pristinamycin



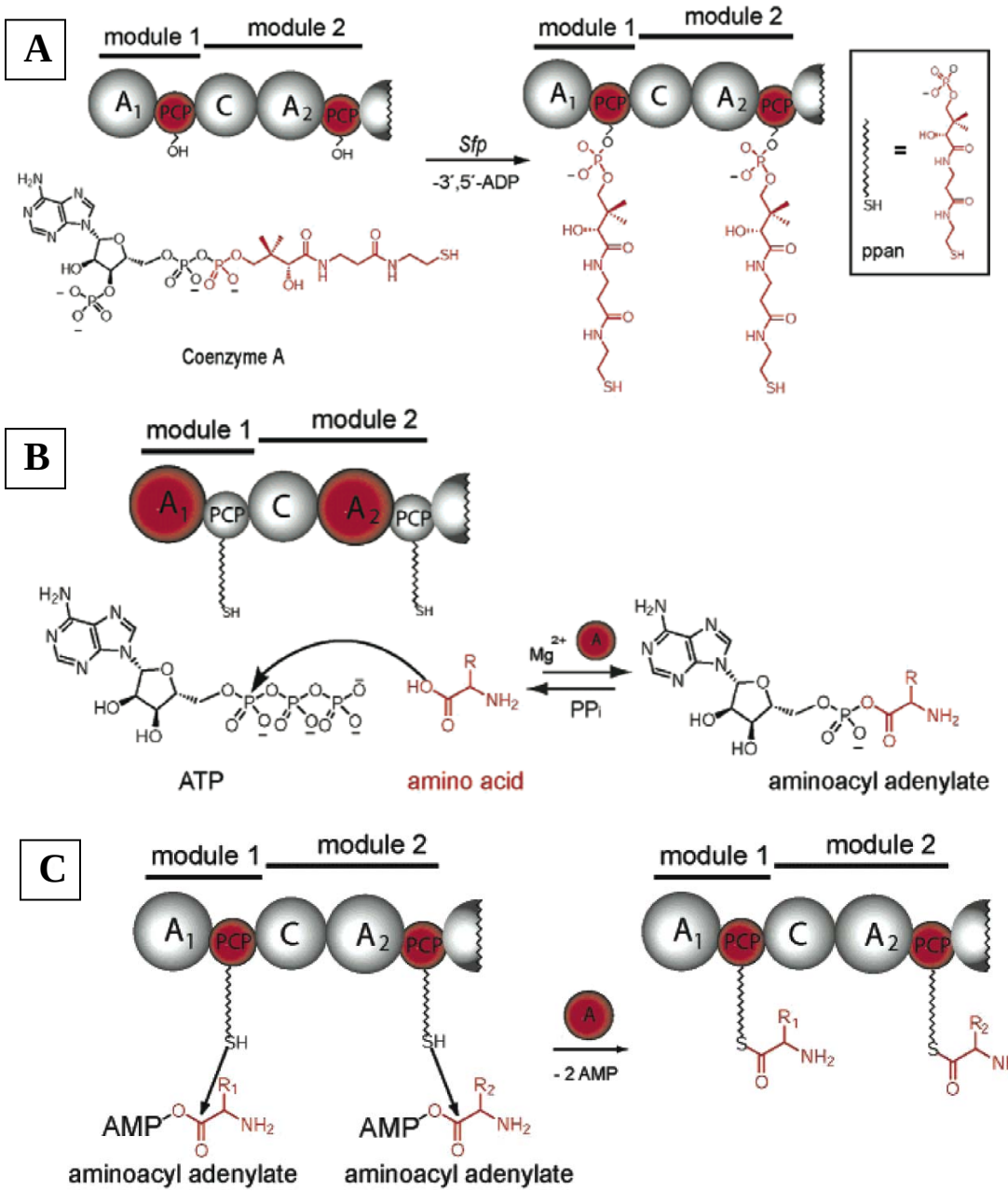
Vancomycin

Non-Ribosomal Peptide Synthesis: Initiation

(**A**) Non-ribosomal peptide bio-synthesis requires proper assembly and modification of a large multienzyme complexes including the addition of the prosthetic moiety 4'-phosphopantetheinyl cofactor (ppan) to a conserved serine residue of chain of **peptidyl carrier protein** (PCP).

(**B**) The first step of peptide assembly is similar to ribosomal protein synthesis: an amino acid is activated by trans-esterification with ATP to afford the corresponding aminoacyl-adenylate. The domains A1 and A2 each bind to a specific amino acid through non-covalent interactions and catalyze the trans-esterification to produce each aminoacyl-adenylate. **"A" = adenylation domain.**

(**C**) Domain A also catalyzes the transfer of each aminoacyl-adenylate onto the free thiol group of PCP-ppan within its module. This establishes a covalent linkage between enzyme and substrate. At this stage the substrate can undergo modifications such as epimerization or N-methylation by neighboring **editing or "E" domains**.

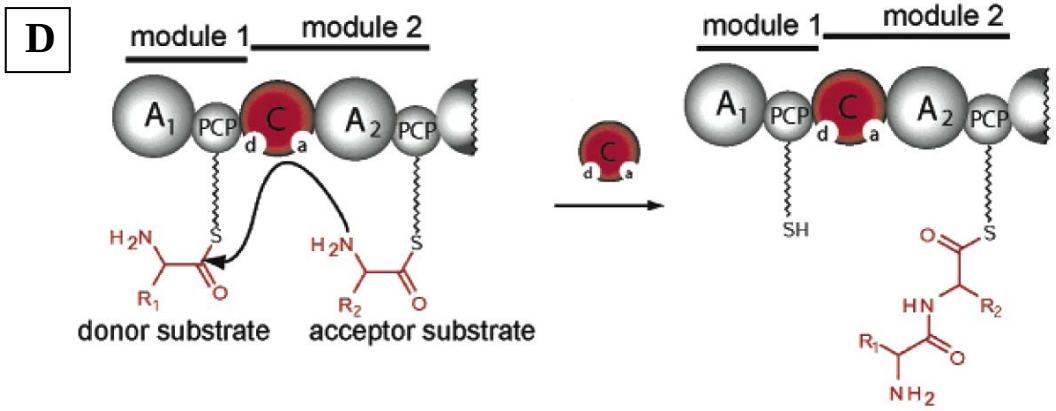


Non-Ribosomal Peptide Synthesis: Elongation

(**D**) Assembly of the product occurs by a series of peptide bond formation steps (elongation) between the downstream building block with its free amine and the carboxy-thioester of the upstream substrate.

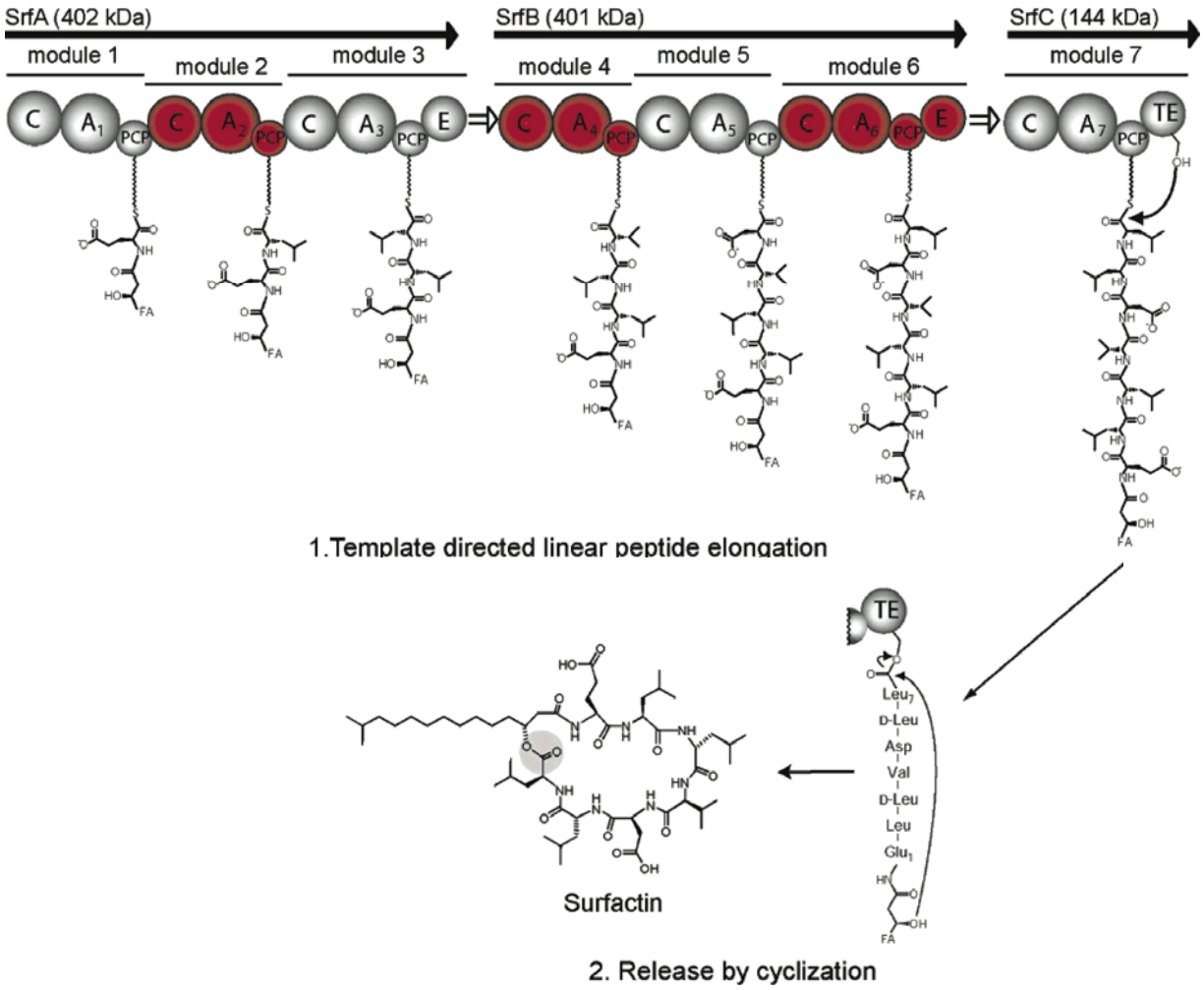
Peptide bond formation is mediated by the **condensation or "C" domain**. The C domain catalyzes the nucleophilic attack of the downstream PCP-bound acceptor amino acid with its free amino group on the activated thioester of the upstream PCP-bound donor.

In contrast to ribosomal protein synthesis with tRNA-bound ester intermediates, nonribosomal peptide synthetases exploit more reactive PCP-thioesters.



Non-Ribosomal Peptide Synthesis

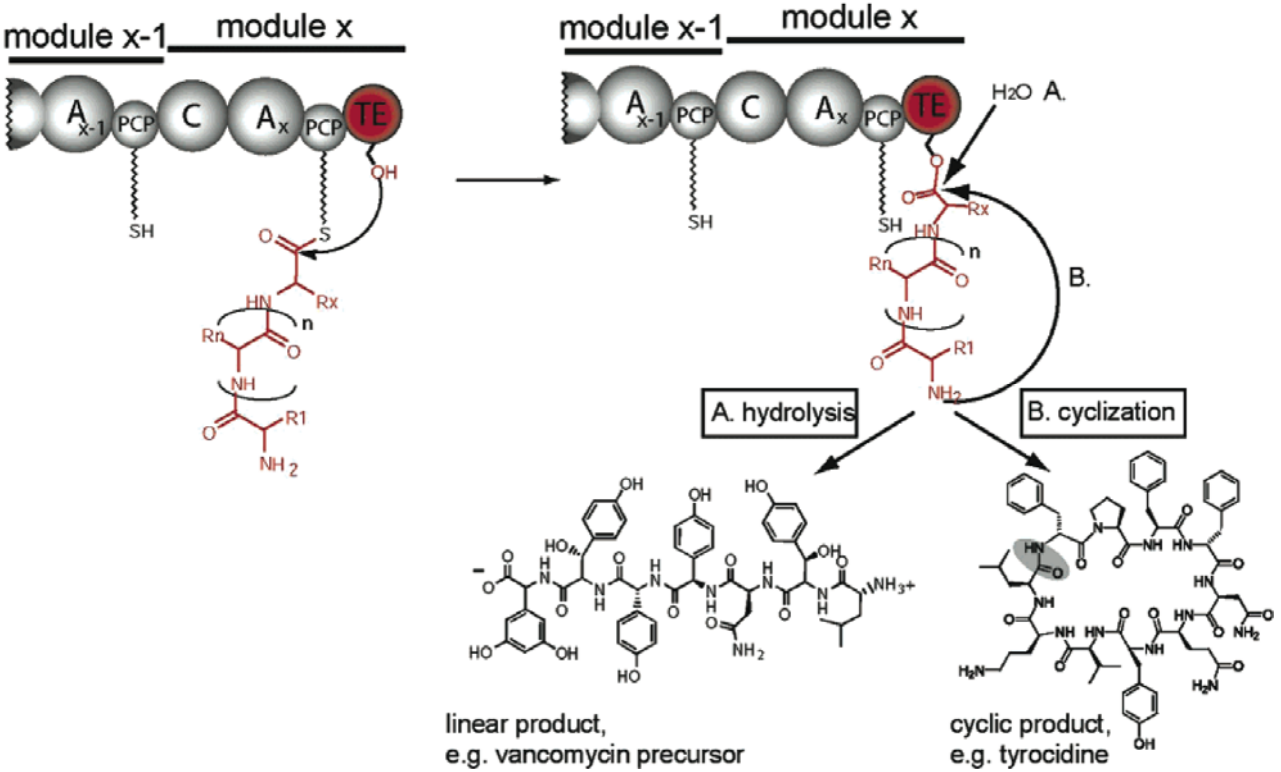
Most biologically produced non-ribosomally produced peptides share a common mode of synthesis, known as the **multienzyme thiotemplate mechanism**. According to this model, peptide bond formation takes place on large multienzyme complexes, which present both a template and the biosynthetic machinery to catalyze all steps of the assembly process. A module is a distinct section of the multienzyme that is responsible for the incorporation of one specific amino acid into the final product. It is further subdivided into a set of domains responsible for substrate recognition, activation, binding, modification, elongation, and release. Domains can be identified at the protein level by characteristic highly conserved sequence motifs. **"TE" = transesterification domain.**



The surfactin assembly line. The multienzyme complex consists of seven modules (grey and red), which are specific for the incorporation of seven amino acids. Twenty-four domains of five different types (C, A, PCP, E, and TE) are responsible for the catalysis of 24 chemical reactions. Twenty-three reactions are required for peptide elongation, while the last domain is unique and required for peptide release by cyclization.

Non-Ribosomal Peptide Synthesis: Termination

Following elongation, the product is released by the action of the **C-terminal thioesterase (TE)** domain which carries out a nucleophilic attack on the PCP-peptidyl thioester to form a covalent acyl-enzyme intermediate. Depending on the particular TE domain, this intermediate can either be released by hydrolysis as a linear acid or by an intramolecular reaction with an internal nucleophile to give a cyclic peptide:

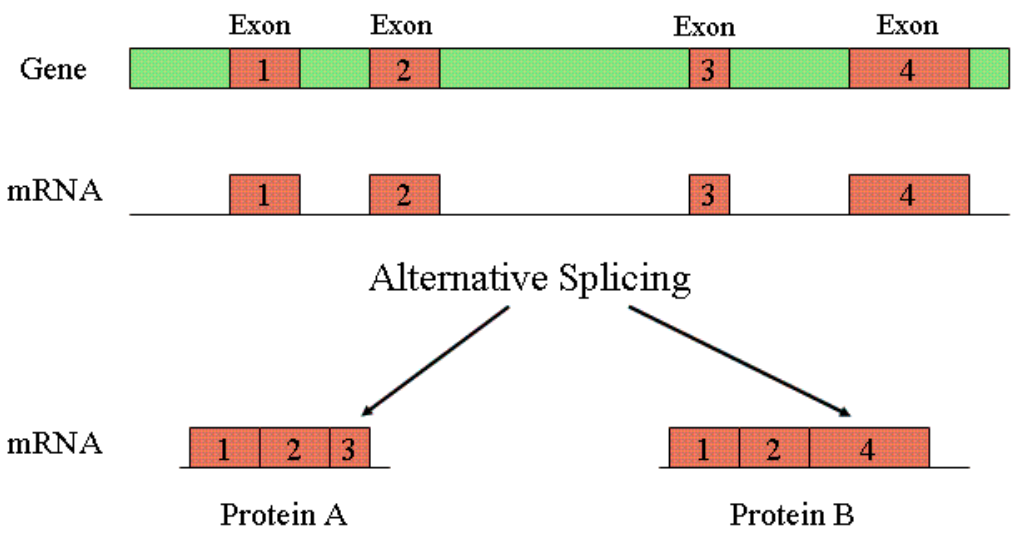


Macrocyclization is a key structural feature of many nonribosomal peptide products. This constrains the flexible peptide chain in a biologically active conformation which can facilitate specific interactions with dedicated cellular targets.

There is another biosynthetic approach to cyclic peptide formation which involves protein splicing of ribosome-produced peptides and proteins.

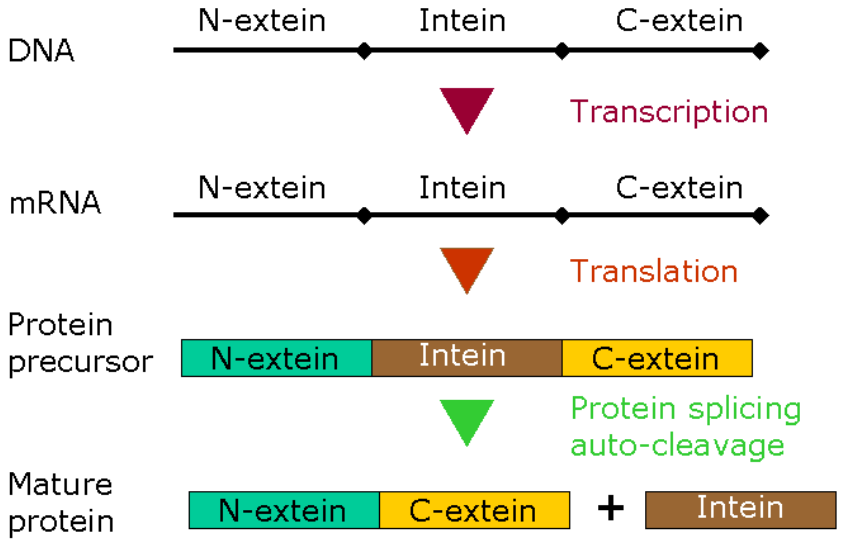
Genome vs. Proteome

Alternative Splicing of RNA



This is called alternative splicing, and can produce different forms of a protein from the same gene. The different forms of the mRNA are called “transcript variants,” “splice variants,” or “isoforms.” The current record-holder for alternative splicing is a Drosophila gene called Dscam which has ~38’000 splice variants. This gene has 95 alternate exons and encodes an axon guidance receptor.

Splicing of Protein*

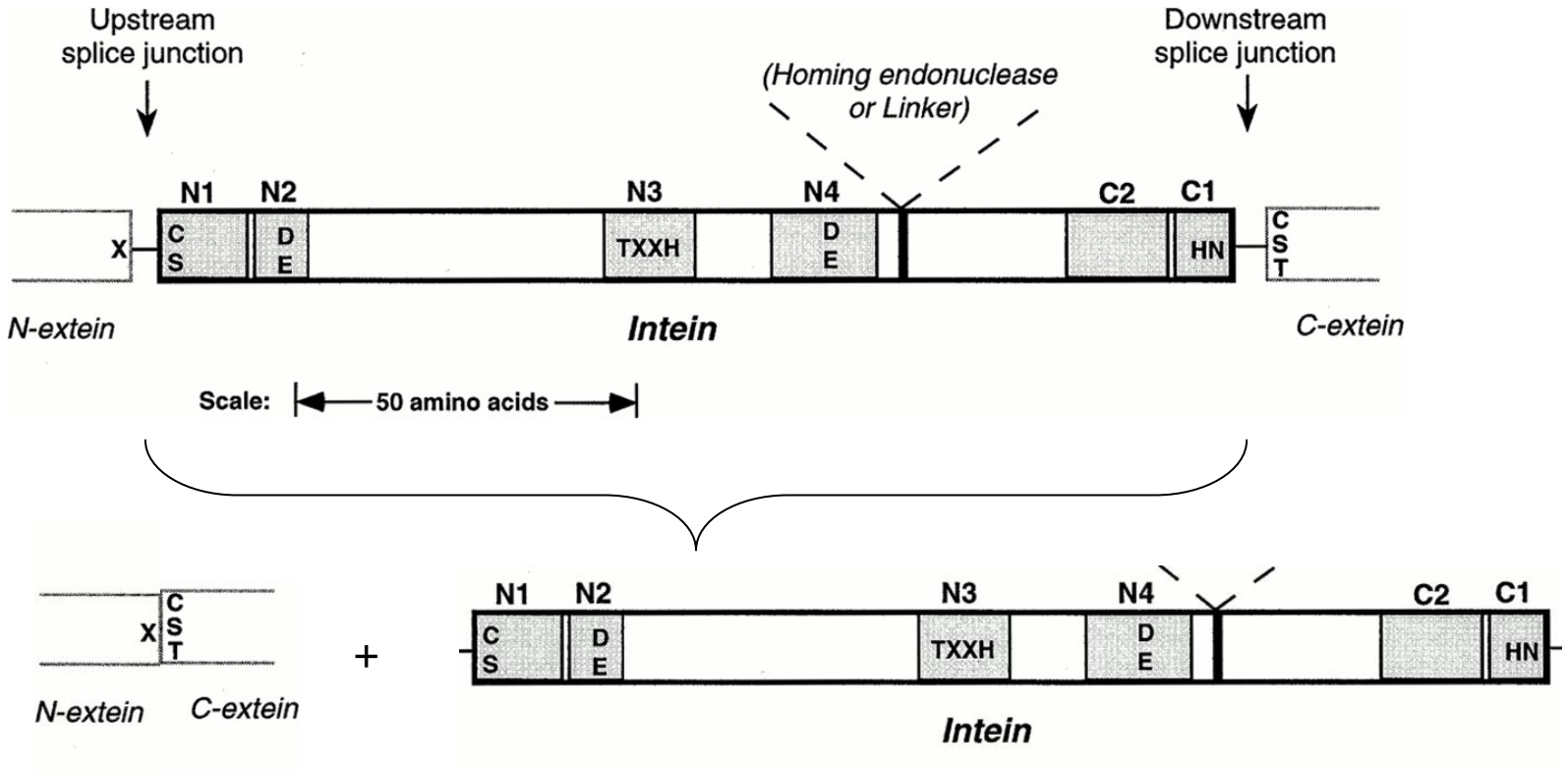


An intein is the protein equivalent of intron. A self-splicing intein catalyzes its own removal from the host protein through a posttranslational process of protein splicing. A mobile intein displays a site-specific endonuclease activity that confers genetic mobility to the intein through intein homing. Inteins can evolve into new structures and new functions, such as split inteins that do trans-splicing.

Annu. Rev. Genet. 2000. 34:61–76

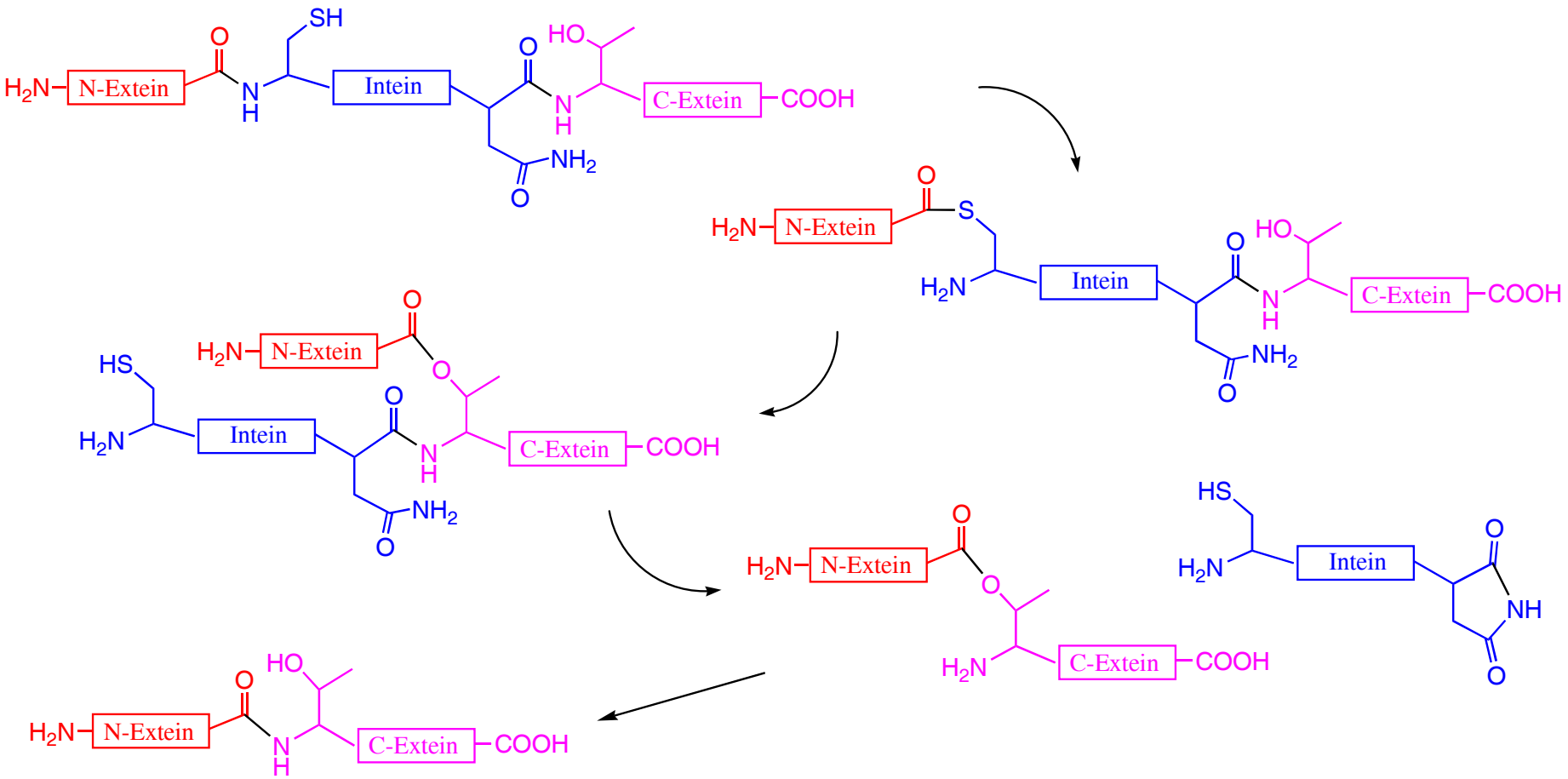
Protein Splicing

Protein splicing is an intramolecular process catalyzed entirely by the intein domain (no cofactors needed). The smallest inteins comprise ca. 150 residues. Around 100 examples have been identified, which are found in all three domains of life (archea, bacteria and eukaryote), suggesting an ancient evolutionary origin. Inteins can be recognized from its amino acid sequence, which has several conserved features:



The **C-terminal residue at either splice junction is always an amino acid with a thiol or alcohol group**. At the intein **C-terminus is a conserved Asn**. Most, but not all, inteins also contain a separate endonucleolytic domain that helps them to maintain their presence by a homing mechanism.

Protein Splicing : Mechanism

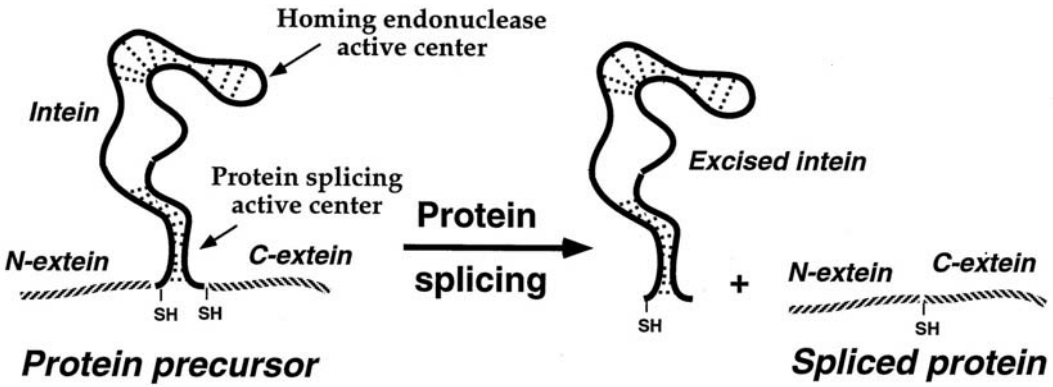


Key insights into possible mechanisms at each step have come from a variety of chemical, biochemical and crystallographic studies. (1) The presence of a thioester as an intermediate after the first rearrangement step was indicated by trapping with NH_2OH. (2) The rearrangement can be blocked by mutation of the C/S/T residue at the N-terminus of the C-extein to A. (3) Replacement of the N-terminal intein residue with an amino acid other than C/S blocks protein splicing.

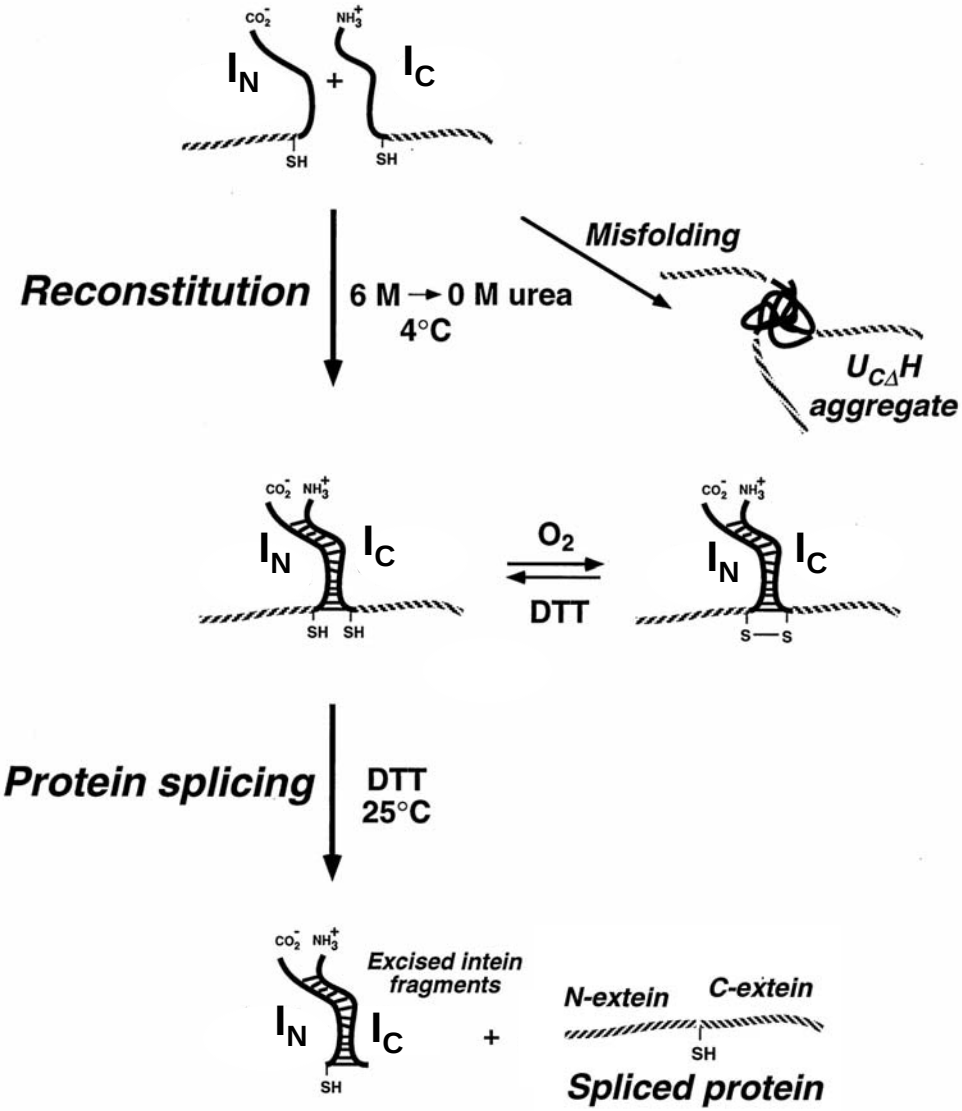
Protein Splicing *In Vitro*

Another important advance came by noting that the endonuclease domain of the intein, accounting for about one-half of its amino acid sequence, can be deleted without affecting the process of protein splicing. This suggested the design of trans-splicing systems in which N-terminal and C-terminal fragments of the intein are co-expressed as separate polypeptides fused to exteins. N- and C-terminal fragments of the *M. tuberculosis* RecA intein, each with about 100 residues, undergo trans-splicing when reconstituted *in vitro*. However, the purified N- and C-intein fragments do not behave like the intact parent protein, and usually require denaturation and renaturation to fold into the active intein. Split inteins can be generated **artifically** at the genetic level, and split inteins also occur **naturally**. Thus providing another potential mechanism for diversity in certain proteomes.

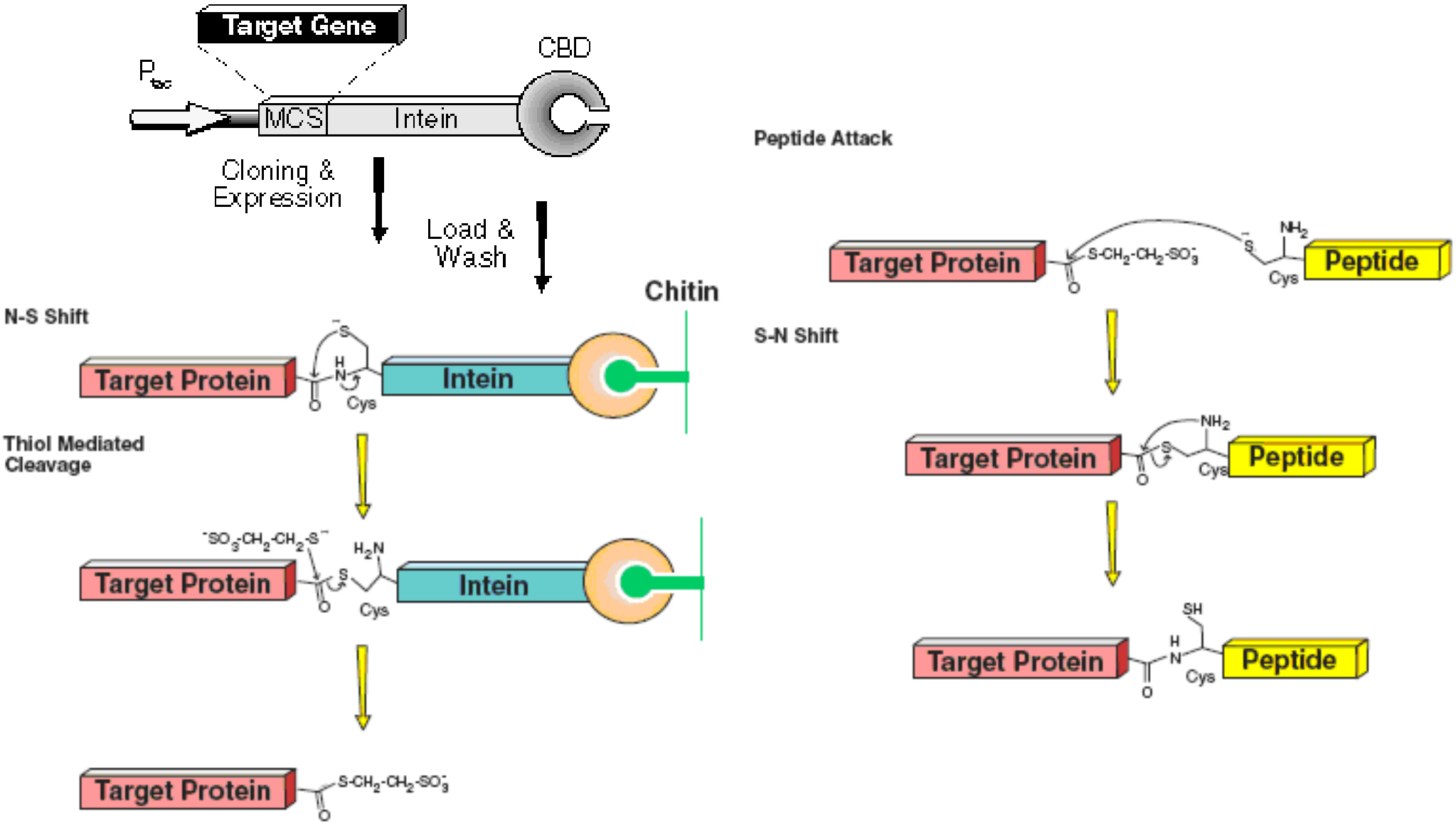
Intramolecular Splicing



Split Intein: Intermolecular Splicing

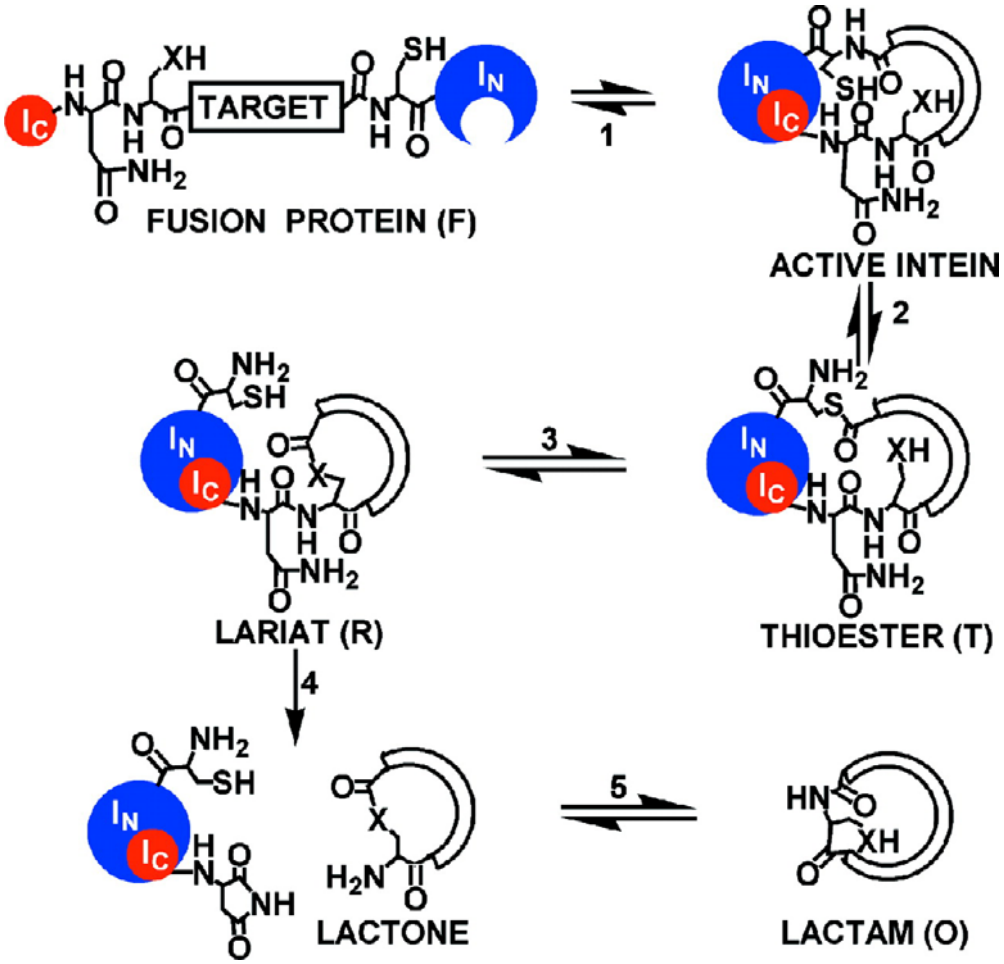


Use of Inteins for Protein Purification and Ligation



Protein Splicing: Cyclic Peptides

Split inteins have also been used to generate libraries of cylic peptides *in vivo*. These cyclic peptides may become a means of generating new drug candidates as the “target” region can be genetically randomized, expressed in cell cultures, and functional cyclic peptides can be identified in the same host organism in which they are expressed.

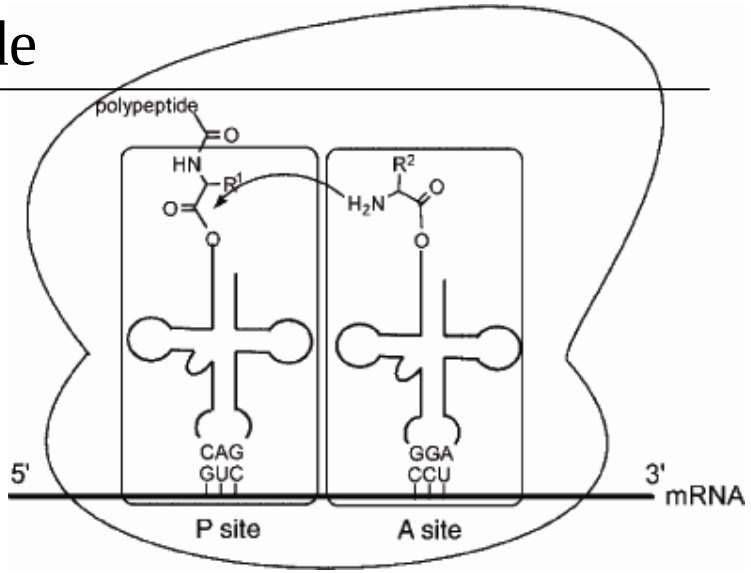


Circular ligation mechanism. An expressed fusion protein (F) folds to form an active protein ligase (1). The enzyme catalyzes an N-to-S acyl shift (2) at the target-IN junction to produce a thioester intermediate (T), which undergoes transesterification (3) with a side-chain nucleophile (X) at the IC-target junction to form a lariat intermediate (R). Asparagine side-chain cyclization (4) liberates the cyclic product as a lactone, and an X-to-N acyl shift (5) generates the thermodynamically favored, lactam product (O) in vivo.

Proc. Natl. Acad. Sci. 1999, 96, 13638-13643.

Expanded Genetic Code

Expanding the genetic code (see *Angew. Chem. Int. Ed. 2005, 44, 34-66.*)



A variety of in vitro (cell-free) methods have been developed to incorporate unnatural amino acids into proteins by using the existing protein biosynthetic machinery of the cell. These methods take advantage of the fact that anticodon-codon recognition between messenger RNA (mRNA) and transfer RNA (tRNA) is largely independent of the structure of the amino acid linked to the 3-terminus of the acceptor stem of the tRNA.

This adaptor hypothesis was elegantly demonstrated by showing that a cysteinyl-tRNACys, when converted into an alanyl-tRNACys by treatment with H_2/Raney nickel, efficiently incorporates alanine into a polypeptide chain in vitro in response to the cysteine codon UGU. These experiments were extended to derivatives of other common amino acids by the chemical modification of enzymatically aminoacylated tRNAs with synthetic reagents and probes. For example, treatment of Lys-tRNALys with N-acetoxysuccinimide affords (N-acetyl-Lys)-tRNALys, which can be used to insert N-acetyl-Lys into proteins in a cell-free rabbit reticulocyte system.

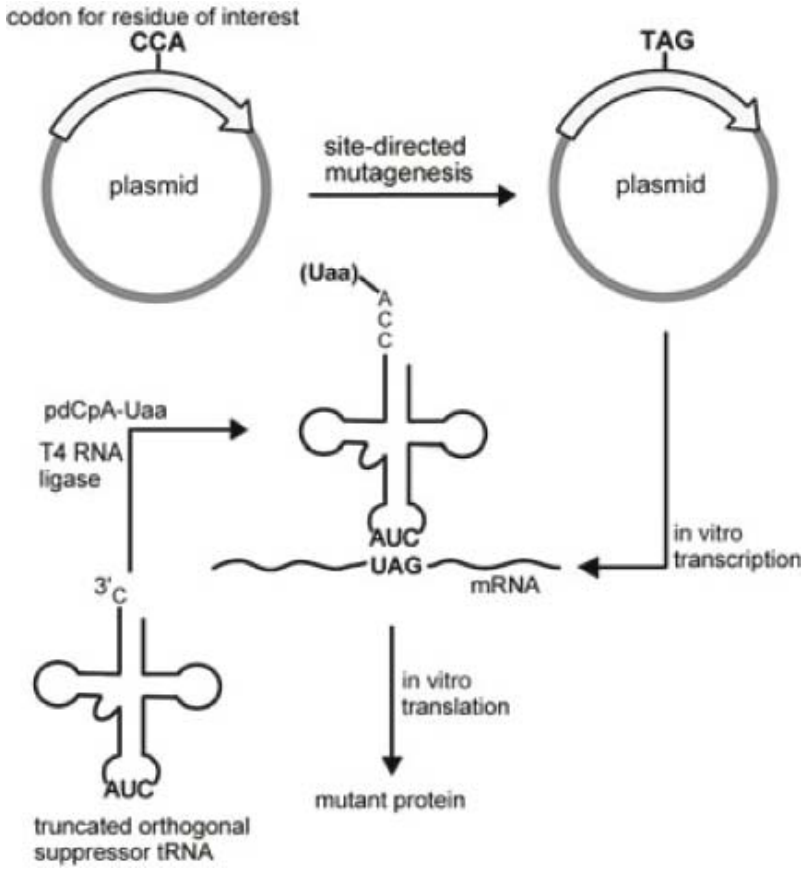
Problems:

- These methods lead to incorporation of the modified amino acid at all sites in a protein specified by the codon for the modified tRNA.
- The derivatized amino acid is also inserted in competition with the underivatized amino acid thus leading to heterogeneous protein products.
- Substitutions are limited to derivatives of the common amino acids, which can be generated under reactive conditions that do not hydrolyze the labile aminoacyl ester linkage between the terminal 3-adenosine group and the amino acid, and do not chemically inactivate the tRNA.

Expanded Genetic Code

A general in vitro method, which allows the site-specific incorporation of a large number of unnatural amino acids into proteins with excellent translational fidelity was reported in 1989. This approach takes advantage of the degeneracy of the three stop codons UAA, UAG, and UGA (termed nonsense codons). These codons do not encode amino acids, but rather signal termination of polypeptide synthesis by binding release factors. Since only one stop codon is required for the termination of protein synthesis, two blank codons exist in the genetic code, which can be used to uniquely specify an unnatural amino acid.

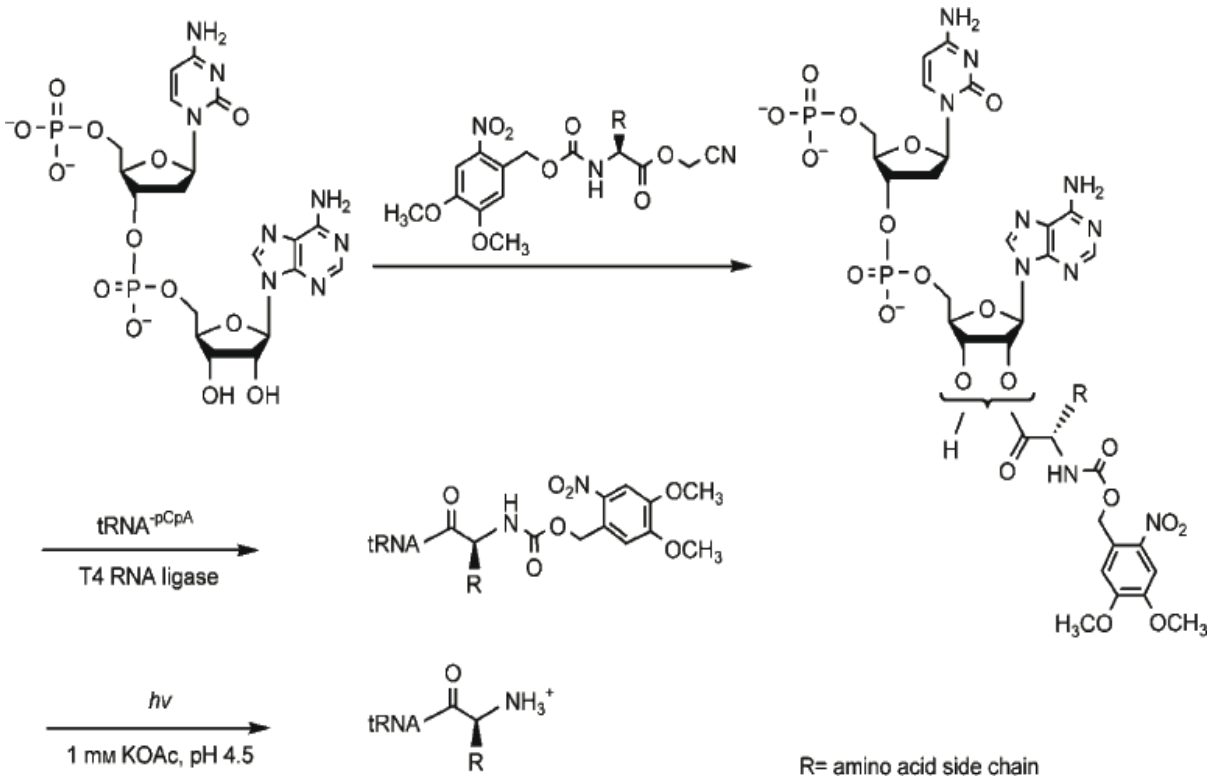
Another requirement for the selective incorporation of unnatural amino acids into proteins is a tRNA that uniquely recognizes this blank codon. Importantly, this tRNA must be orthogonal to the endogenous aminoacyl-tRNA synthetases of the host organism, that is, it must not be a substrate for any of the aminoacyl-tRNA synthetases present. If the suppressor tRNA were recognized by an endogenous aminoacyl-tRNA synthetase, the tRNA could be subject to proofreading (deacylation of the noncognate amino acid) and/or re-aminoacylation with the cognate amino acid. This would result in either low suppression efficiency or the incorporation of a common amino acid in competition with the desired unnatural amino acid in response to the nonsense codon. An orthogonal amber suppressor tRNA derived from yeast phenylalanine-tRNA was constructed for use in an E. coli in vitro protein transcription-translation system. In this tRNA nucleotides 34-37 in the anticodon loop were replaced by 5'-CUAA-3' using run-off transcription:



Expanded Genetic Code: Aminoacyl tRNA Synthesis

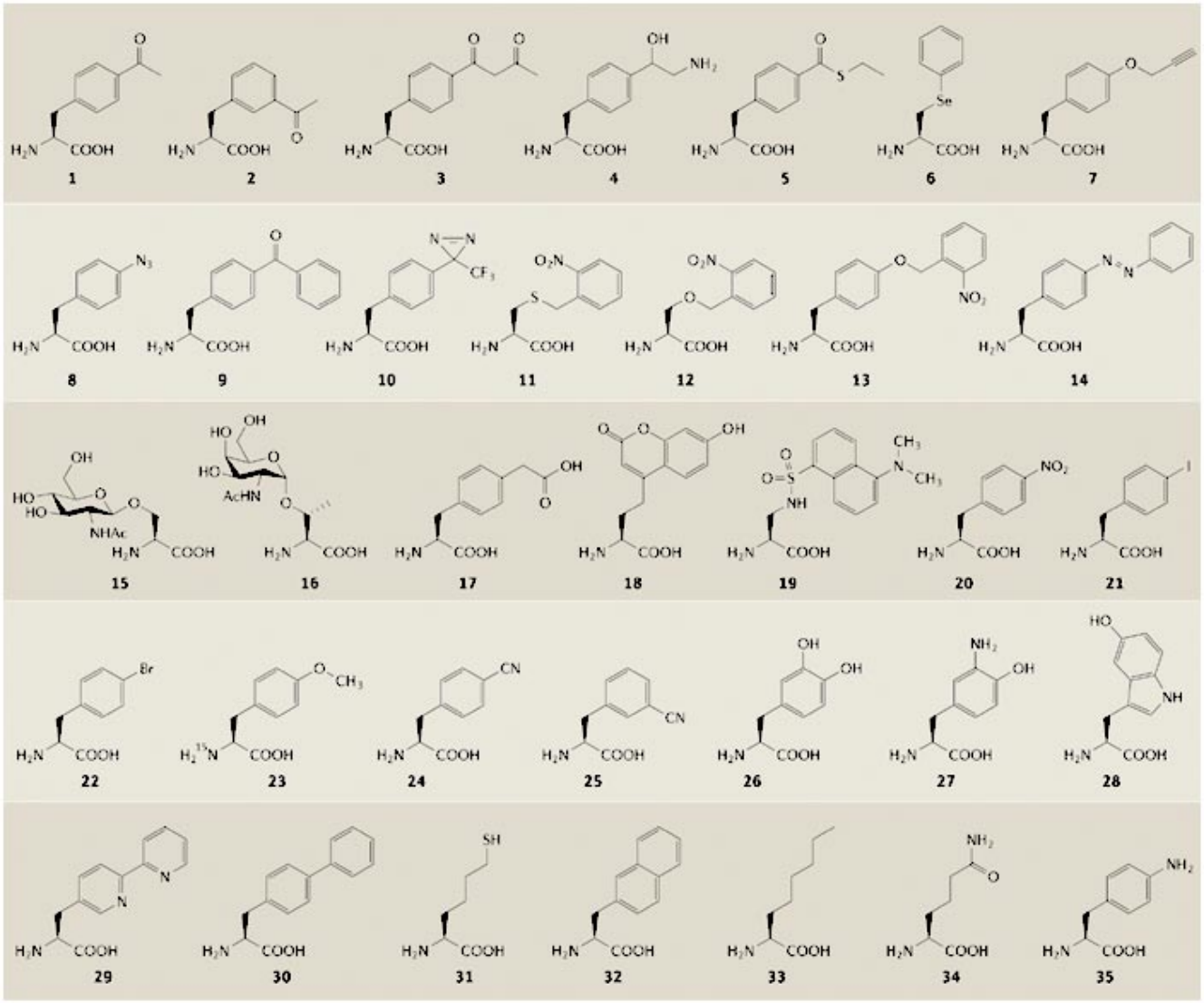
Extension of this approach to a larger variety of amino acids required more general methods for the selective aminoacylation of tRNAs. Direct chemical acylation of the tRNA is not practical because of the large number of reactive sites in the tRNA. To efficiently aminoacylate the orthogonal tRNA, the cyanomethyl ester of an N-protected amino acid is used to monoacylate the 2,3-hydroxy groups of a dinucleotide to give the desired aminoacylated product in high yield, thereby obviating the need to protect and deprotect the dinucleotide:

Replacement of cytidine with deoxycytidine in pCpA simplified the synthesis and eliminated another reactive 2-OH group without affecting biological activity. The α-amino group of the amino acid is protected as a nitroveratryloxy carbamate, ester. This protecting group can be removed photochemically from the intact aminoacyl-tRNA (after ligation) in high yield under mildly acidic conditions. This aminoacylation protocol is relatively straightforward, proceeds in high yield (both the aminoacylation and ligation reactions), and has been used to incorporate a wide variety of unnatural amino acids into proteins.



Expanded Genetic Code

Since the first demonstration of this general approach (Science 1989, 244, 182), over 50 different unnatural amino acids have been site-specifically incorporated into proteins including:



Reactive side chains

Photocross-linking residues

“Caged” side chains

Glycosylated residues

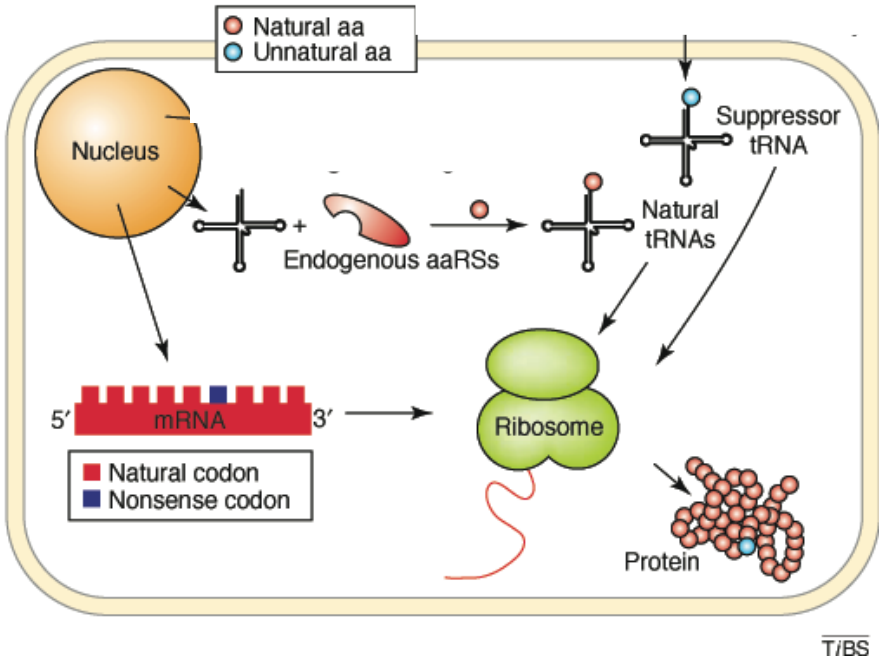
Fluorescent probes

Metal chelating

Expanded Genetic Code

Some natural aminoacyl-tRNAs are tolerated by the ribosome, others are not.

In addition, the unnatural aminoacyl-tRNA must be efficiently transported into the cytoplasm when it is added to the growth medium or biosynthesized by the host, and it must be stable in the presence of endogenous metabolic enzymes:



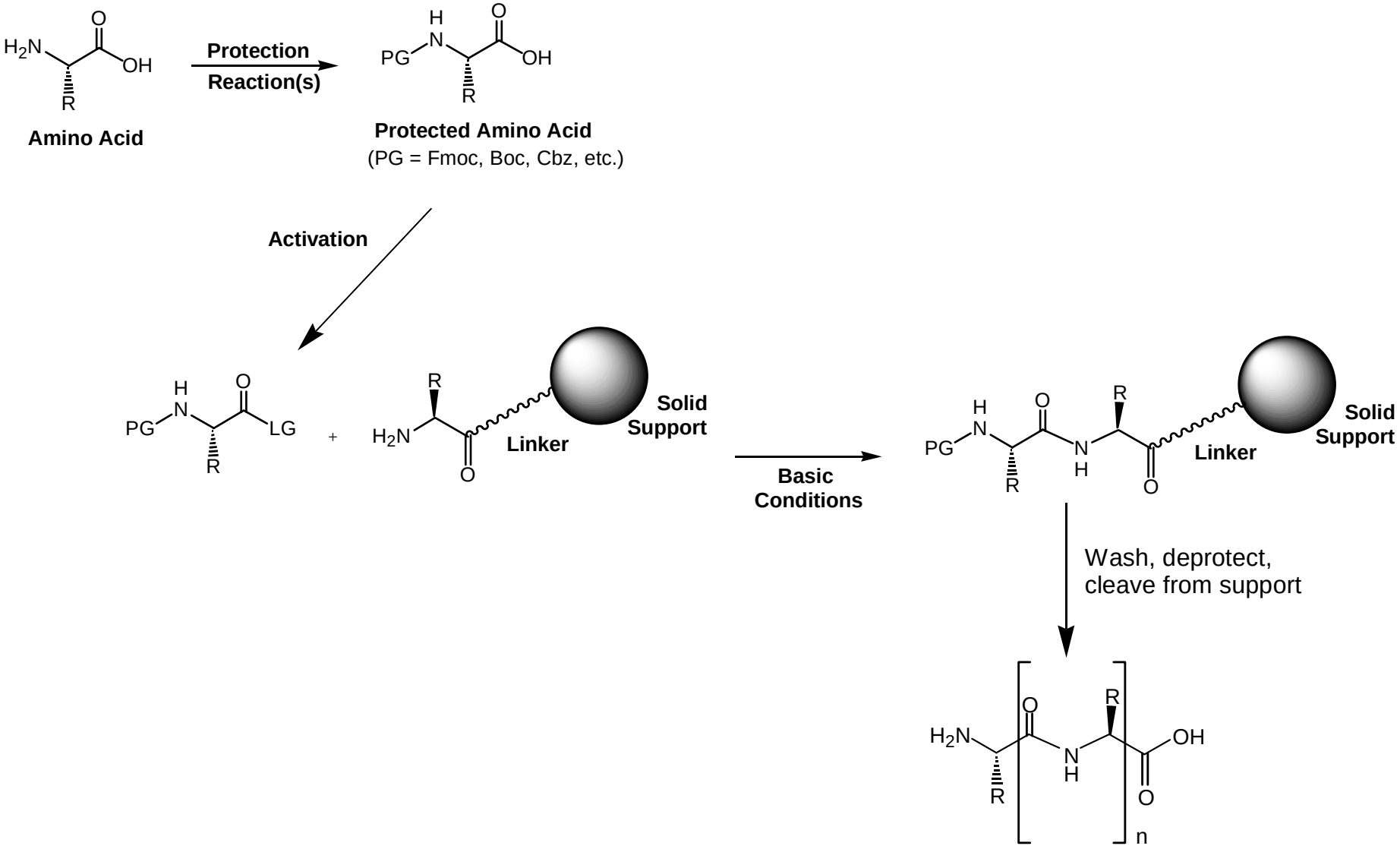
Recently, Peter Schultz's Laboratory has used genetic manipulation to engineer a bacteria able to synthesize p-aminophenylalanine (**35** previous slide) and incorporate it into proteins. To do this, a mutant aminoacyl-tRNA synthetase capable of loading this amino acid onto a mutant suppressor tRNA was evolved.

"We have generated a completely autonomous bacterium with a 21 amino acid genetic code. This bacterium can biosynthesize a nonstandard amino acid p-aminophenylalanine from basic carbon sources and incorporate this amino acid into proteins. Denaturing gel electrophoresis and mass spectrometric analysis show that p-aminophenylalanine is incorporated into myoglobin with fidelity and efficiency rivaling those of the common 20 amino acids. **This and other such organisms may provide an opportunity to examine the evolutionary consequences of adding new amino acids to the genetic repertoire, as well as generate proteins with new or enhanced biological functions**."

J Am Chem Soc. 2003 Jan 29;125(4):935-9.

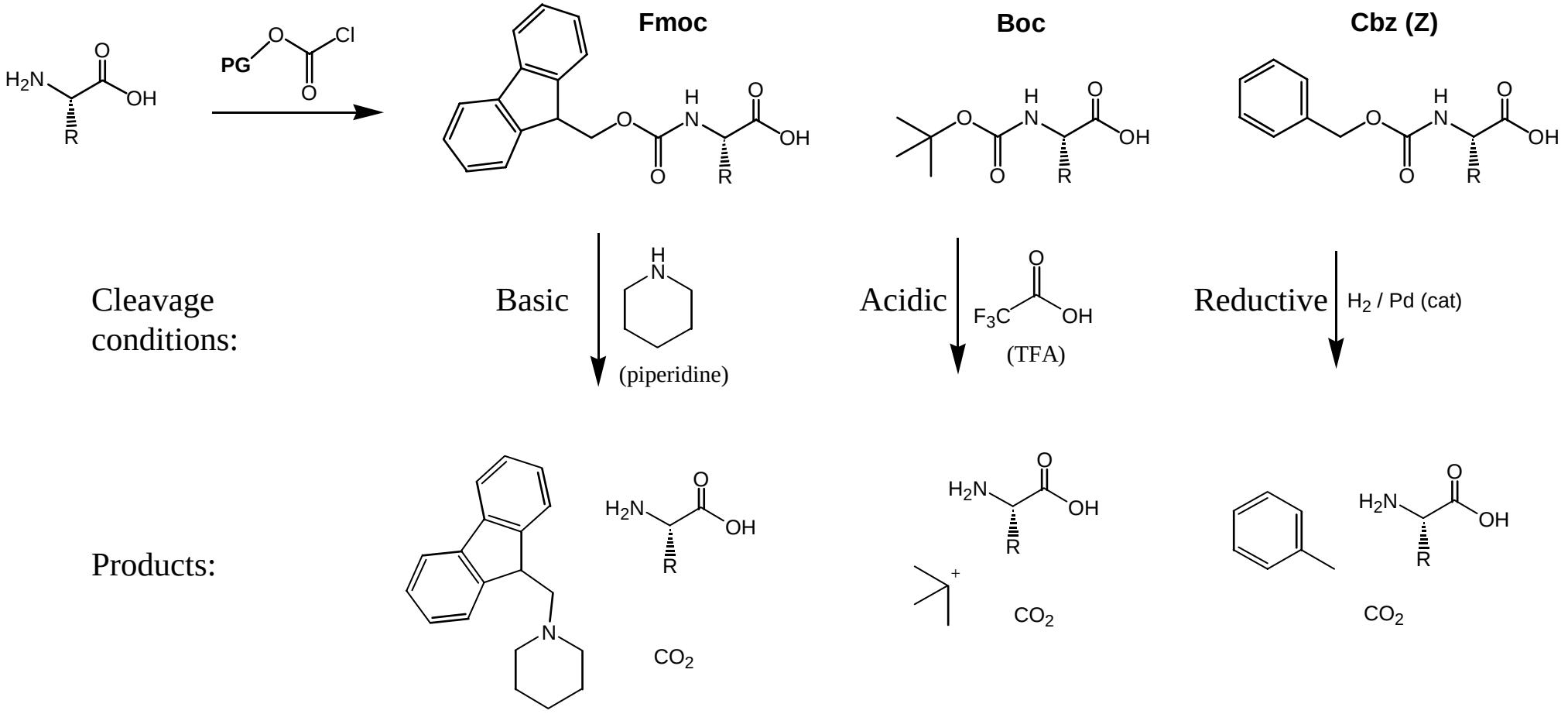
Chemical Synthesis of Peptides and Proteins: General Approach

Chemical synthesis remains the most flexible approach to the synthesis of modified peptides and proteins. In theory, R = “any” natural or unnatural side chain:



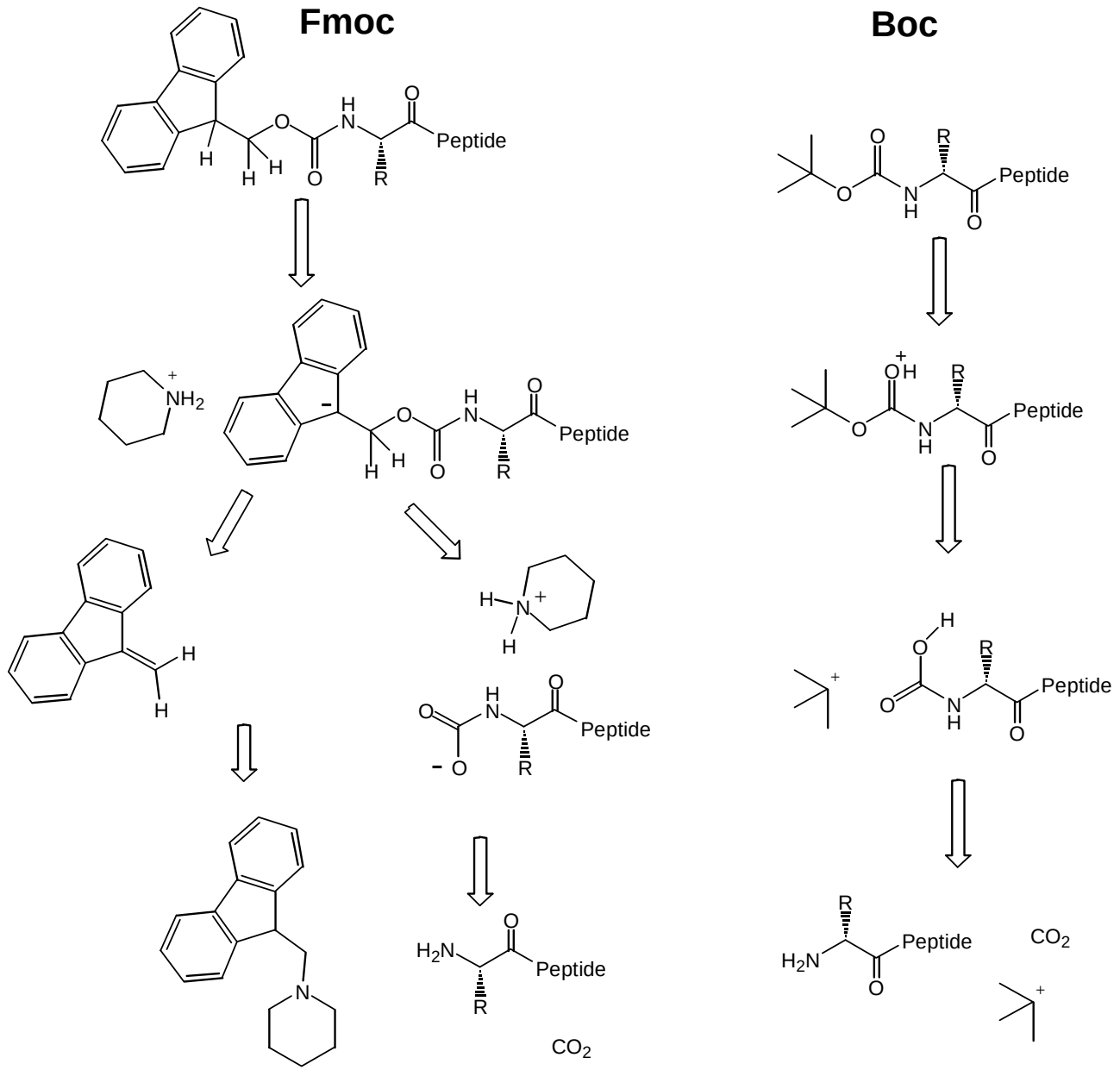
Chemical Synthesis: Protection for the N(α) Amino

Carbamates

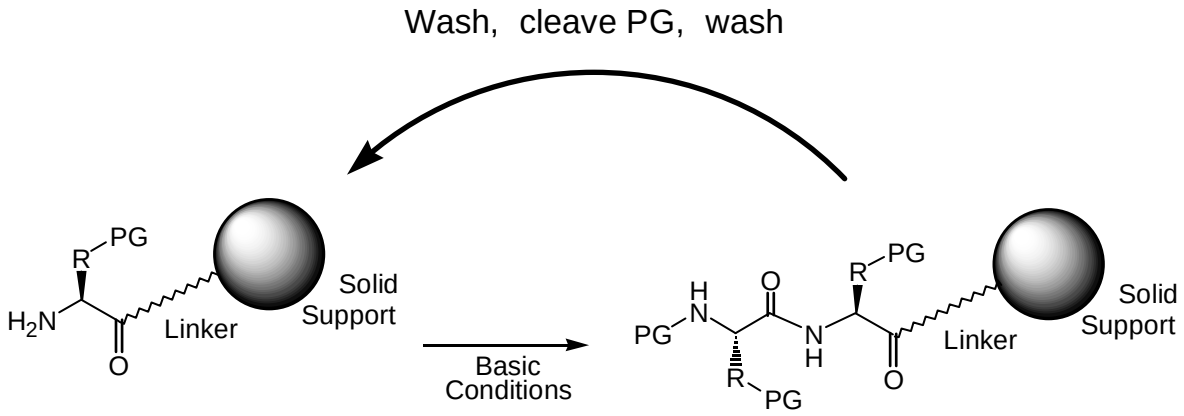


Chemical Synthesis: Protection for the N(α) Amino

Cleavage mechanisms:

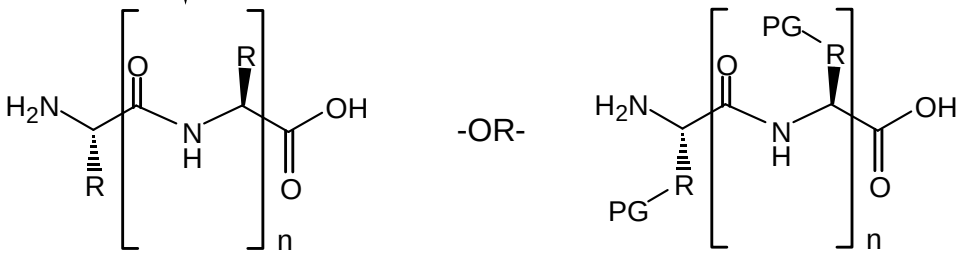


Chemical Synthesis: Protection for Side Chains



N(α) PG cleavage must not affect the PG on the side chains, i.e. they must be "orthogonal."

N(α) PG cleavage must not cleave the peptide from the support.



Cleavage of the protecting groups of the side chains can occur under the same conditions as cleavage from the linker.

-OR- these conditions can also be orthogonal, thus giving a protected peptide.

Chemical Synthesis: Protection for Side Chains

Which side chains need protection?

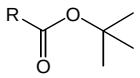
-Nucleophilic under basic conditions

-Can react with activating agent

-Facilitate racimization or other side reactions

What are some common side chain protecting groups?

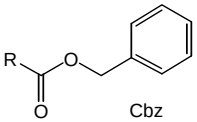
Fmoc



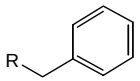
BOC
(Lys, Trp)

OtBu
(Asp, Glu, Tyr)

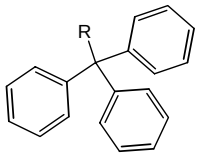
BOC



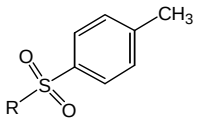
Cbz
(Lys, Trp)



Bzl
(Asp, Glu, Ser, Thr)

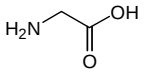


Trityl
(Cys, Gln, Asn, His)

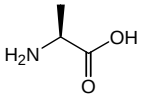


Tos
(Arg, His)

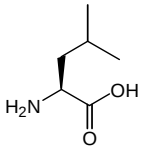
Aliphatic



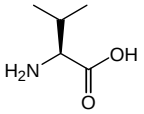
G - glycine (Gly)



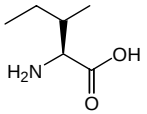
A - alanine (Ala)



L - leucine (Leu)

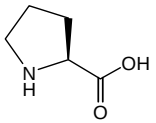


V - valine (Val)



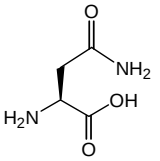
I - isoleucine (Ile)

Cyclic

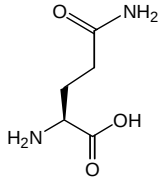


P - proline (Pro)

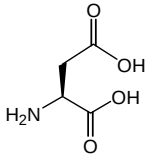
Amide and Acidic



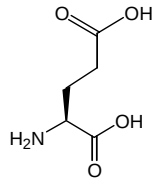
N - asparagine (Asn)



Q - glutamine (Gln)

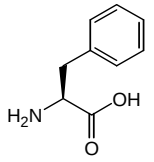


D - aspartic acid (Asp)

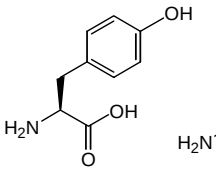


E - glutamic acid (Glu)

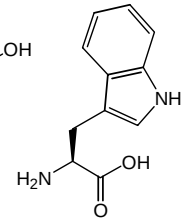
Aromatic



P - phenylalanine (Phe)

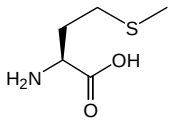


Y - tyrosine (Tyr)

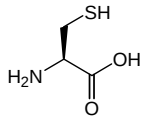


W - tryptophane (Trp)

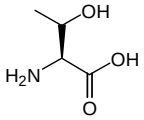
Hydroxyl or Sulfur



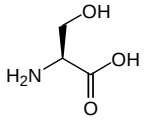
M - methionine (Met)



C - cysteine (Cys)

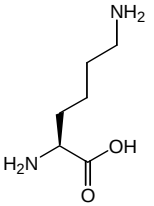


T - threonine (Thr)

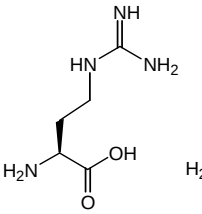


S - serine (Ser)

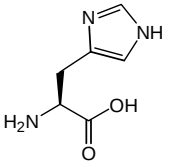
Basic



K - lysine (Lys)



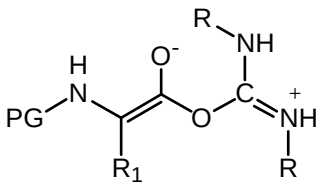
R - arginine (Arg)

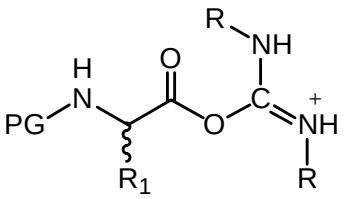


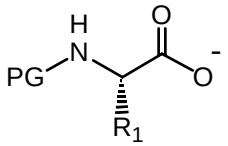
H - histidine (His)

Chemical Synthesis of Peptides and Proteins: Activation

Problem w/ epimerization (also acyl halides):

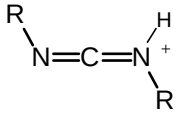






Protected Amino Acid

Carbodiimdes

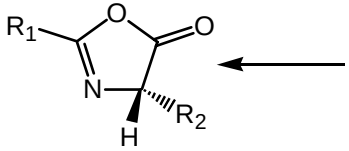


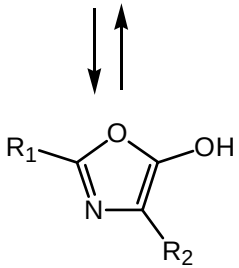
R = cylclohexyl (DCC)
R= isopropyl (DIC)

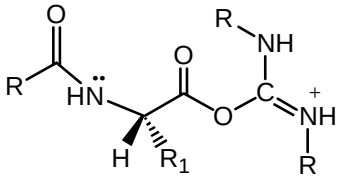
- H^+

+ H^+

Oxazolone (azlactone) formation:

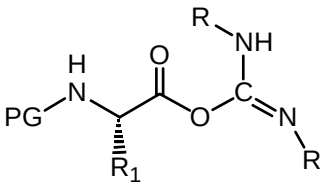






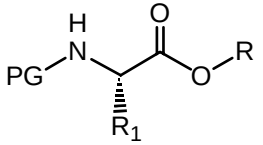
- H^+

+ H^+



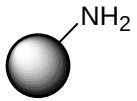
One way to help:

Add large excess:
R-OH

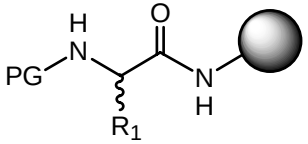


Side
Reaction(s)

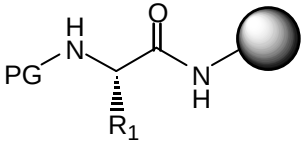
Reaction



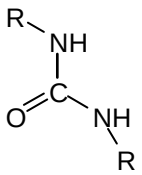
Reaction



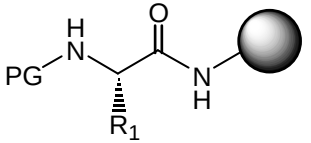
Amide



Amide



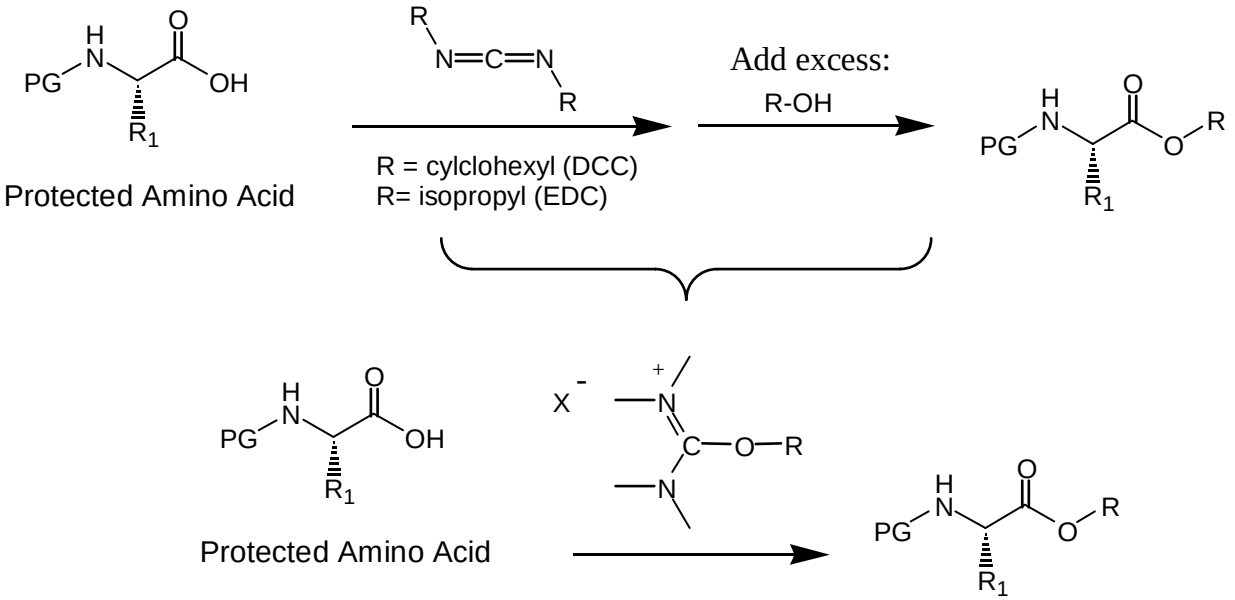
Urea



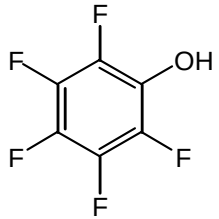
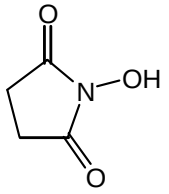
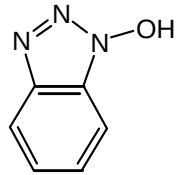
Amide

R-OH

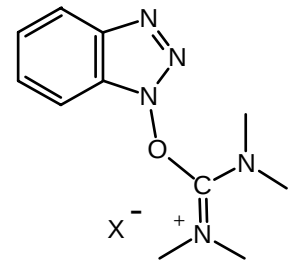
Chemical Synthesis of Peptides and Proteins: Activation



Common R-OH

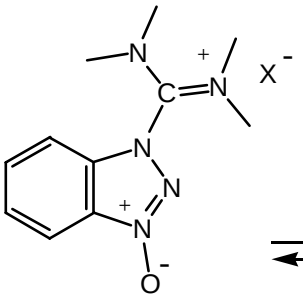
Pentafluoro-phenol	N-hydroxy succinimide	Hydroxyl-benztriazole (HOBT)
		
Requires step-wise addition, limited solubility.	Less reactive ester, but water stable.	Highly reactive ester, little if any epimerization.

Most common activating agents:

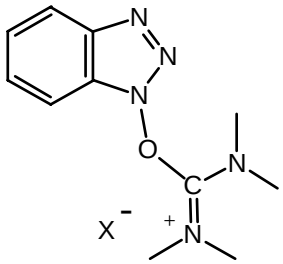


X = PF_6 (HBTU)
X = BF_4 (TBTU)

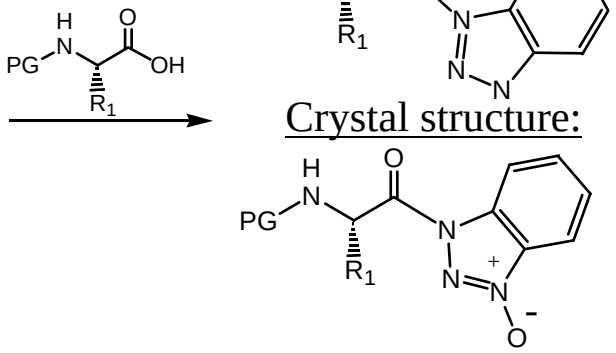
Crystal structure



In solution



In solution:



Crystal structure: