Option Subject

Advanced Chemical Biology

3rd year Hilary Term 8 Lectures

BGD and CJS

Segment 2: Making and Modifying Proteins

Ben.Davis@chem.ox.ac.uk
http://www.chem.ox.ac.uk/researchguide/bgdavis.html

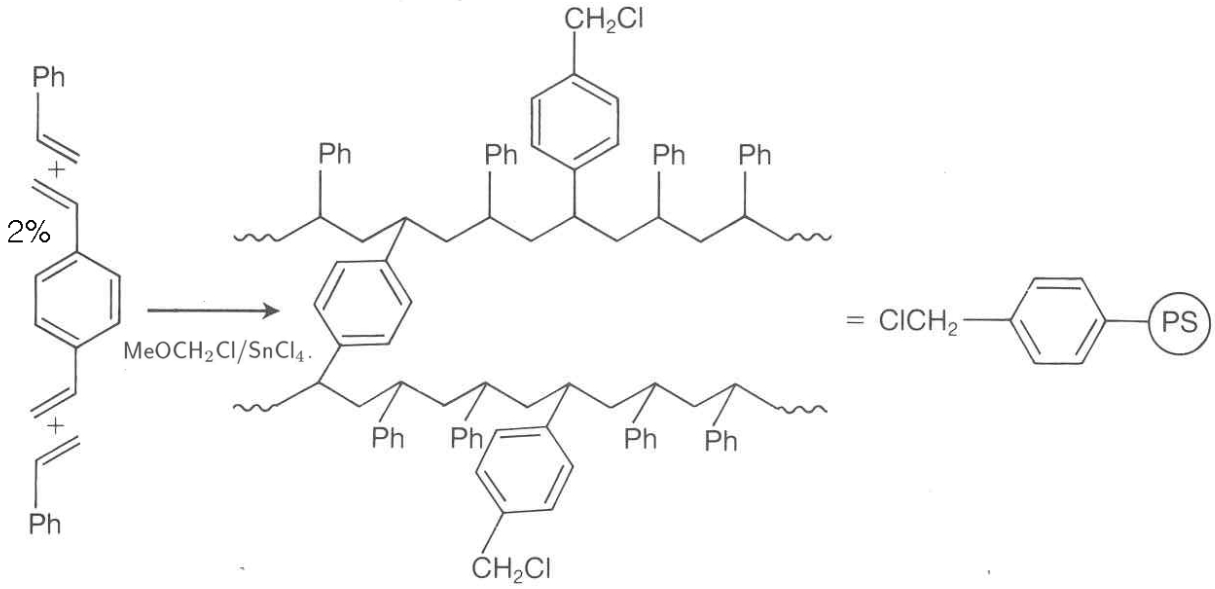
Lecture 3: Making Proteins. Bigger Peptides. Solid-phase Methods. Segment Assembly. Natural Peptide Synthesis. Ribosome Mechanism. The Proteome.

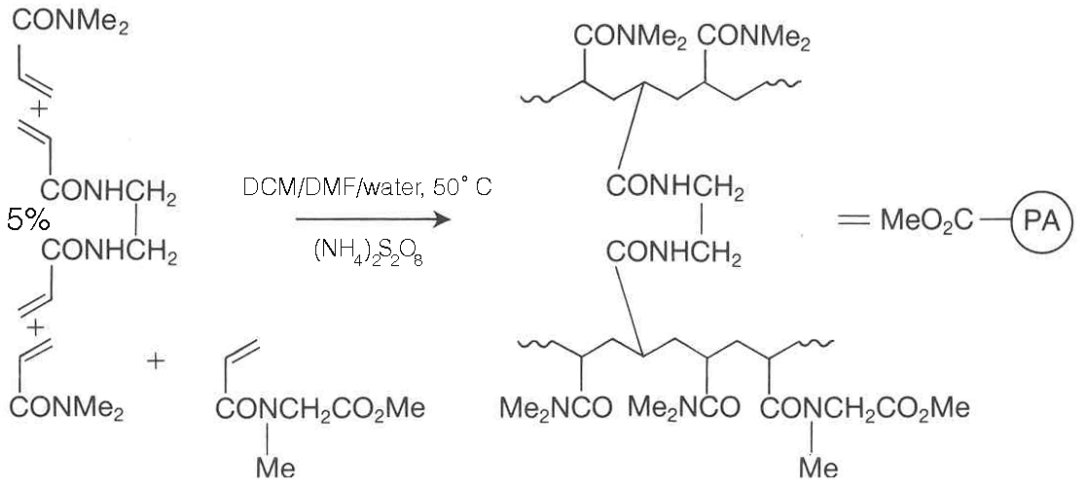
Lecture 4: Other Natural Acyl Transfer Mechanisms. Proteases. Sequencing for Proteomics. Inteins and Native Chemical Ligation. Modified Proteins. Post-translational Modifications. Chemical Methods of Modifying Proteins.

Good books:

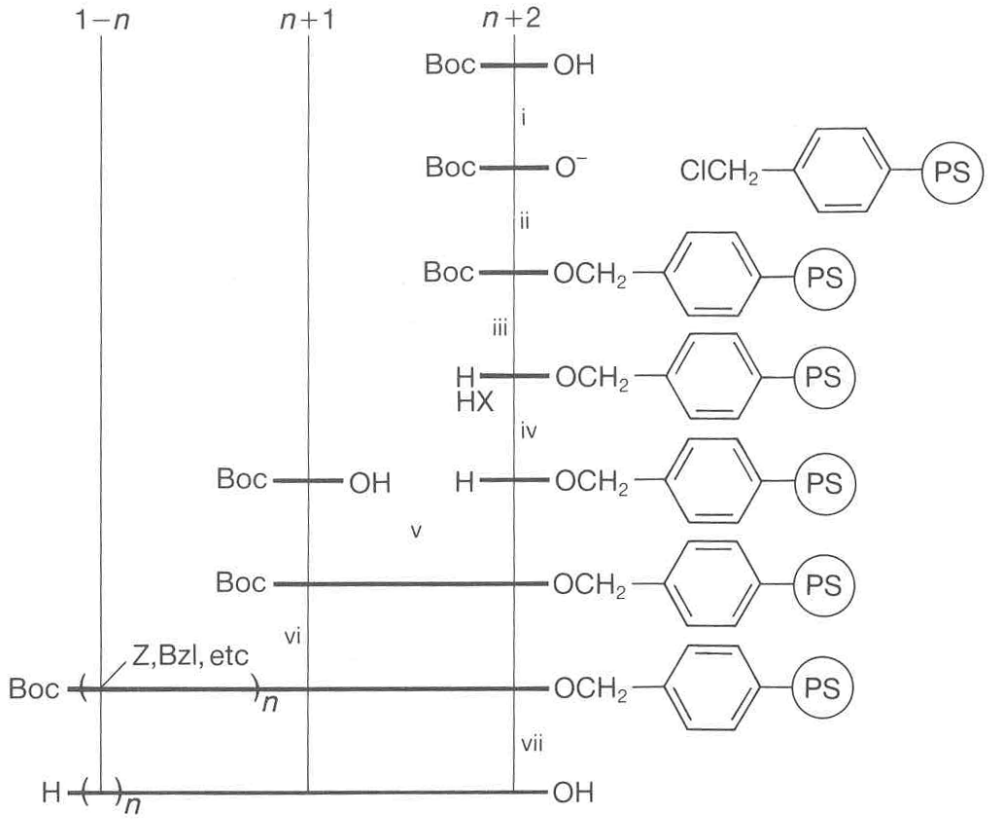
J. Jones, *Amino Acid and Peptide Synthesis*, OCP 7, 2nd edition, 2002
J. Jones, *The Chemical Synthesis of Peptides*, Clarendon Press, 1994
C.T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts & Co, 2005

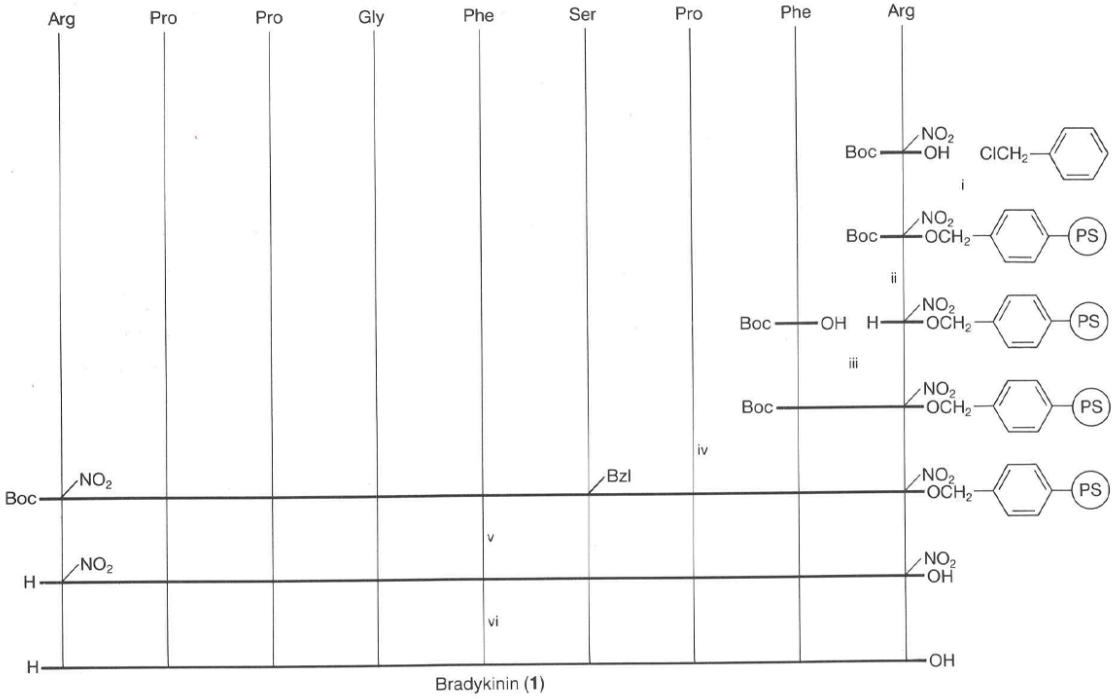
Solid Supports





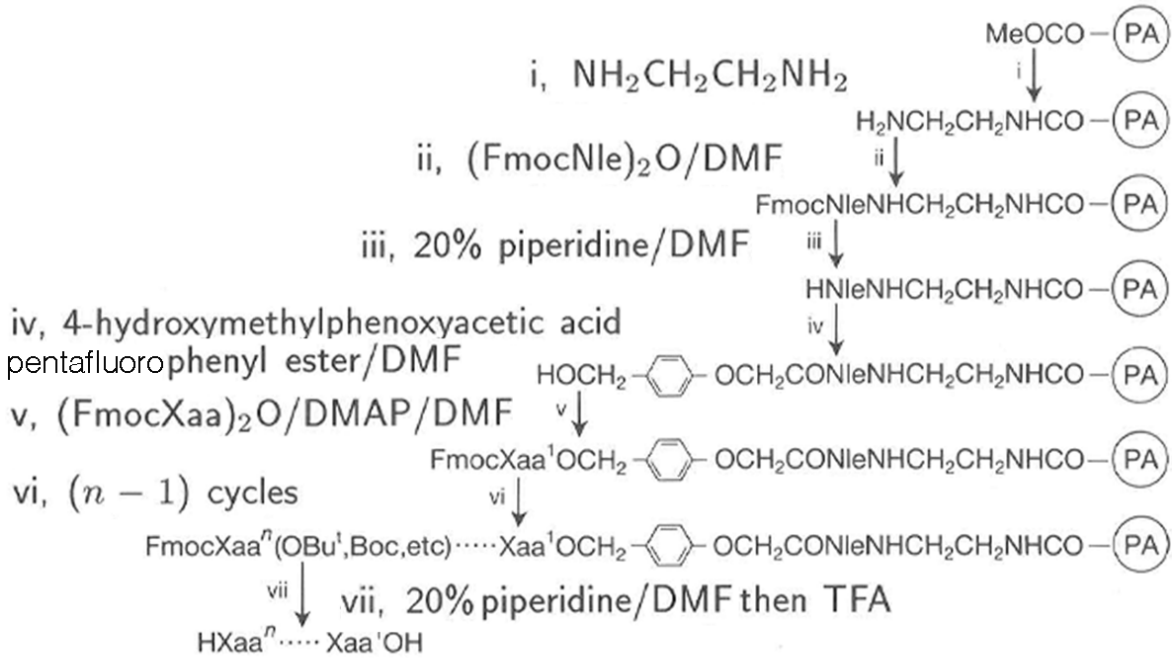
Merrifield Method



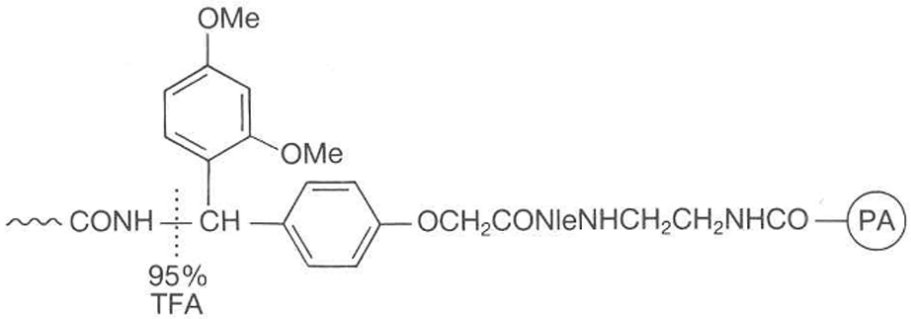


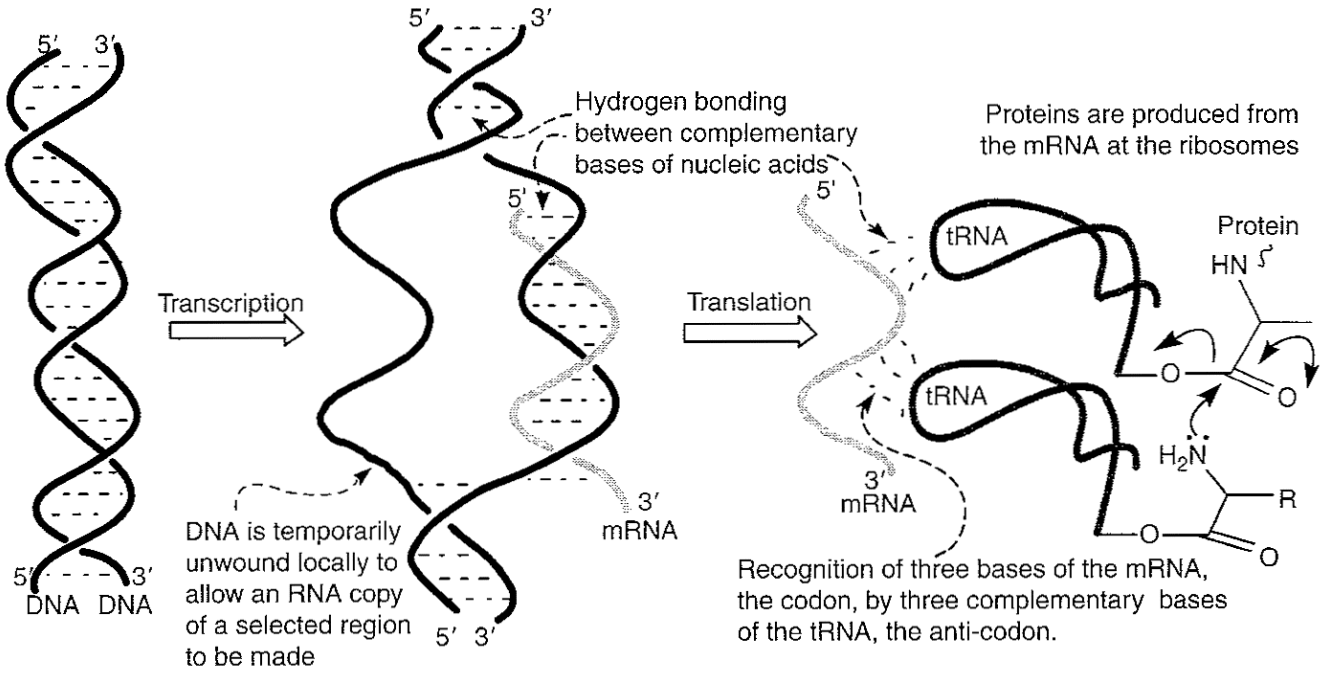
Scheme Conditions: i, Et_3N/EtOH/80 °C/48 h; ii, HCl/AcOH, then Et_3N/DMF; iii, DCCI/DMF; iv, seven similar cycles of deprotection, neutralization, and coupling with the appropriate Boc amino acids; v, HBr/TFA; vi, H_2/Pd(C)/MeOH, followed by ion exchange chromatography purification.

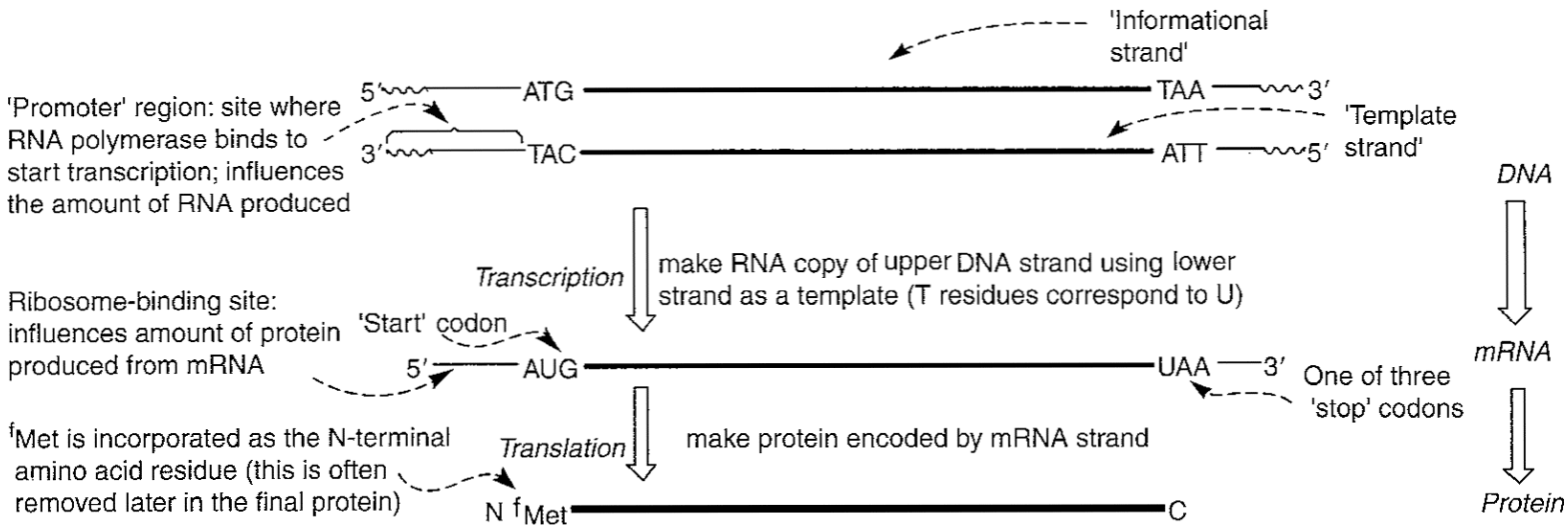
Sheppard Appoach



or TFA




5′ 3′
5′ 3′
DNA DNA
Transcription
5′ 3′
Hydrogen bonding
between complementary
bases of nucleic acids
5′
3′
mRNA
DNA is temporarily
unwound locally to
allow an RNA copy
of a selected region
to be made
5′ 3′
Translation
Proteins are produced from
the mRNA at the ribosomes
5′
tRNA
tRNA
3′
mRNA
Protein
HN
O
O
H2N
R
O
O
Recognition of three bases of the mRNA,
the codon, by three complementary bases
of the tRNA, the anti-codon.

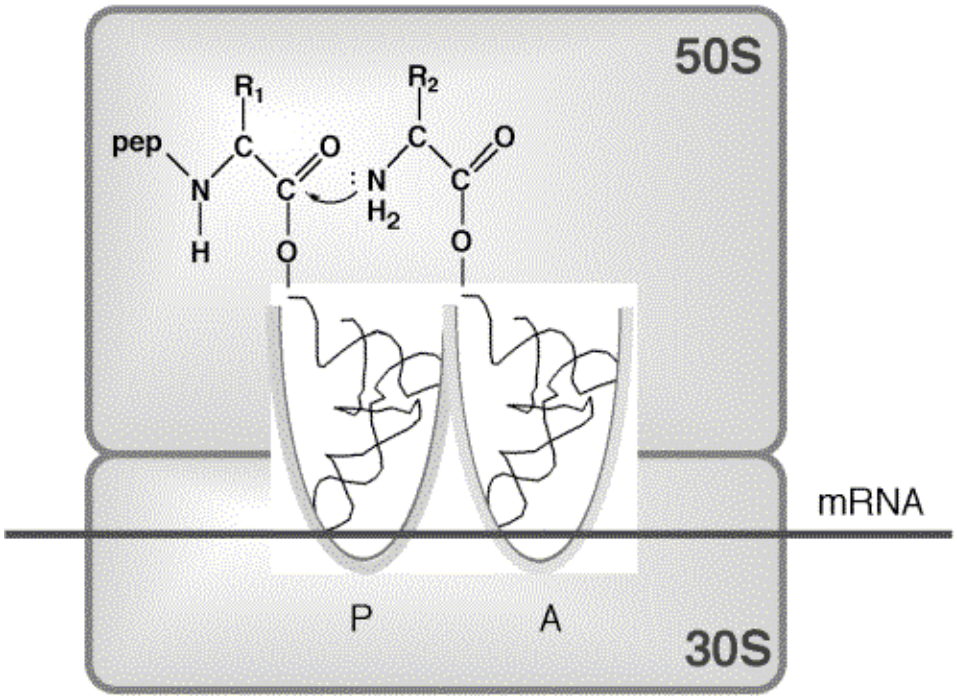

'Informational
strand'
'Promoter' region: site where
RNA polymerase binds to
start transcription; influences
the amount of RNA produced
5′
ATG
TAA
3′
3′
TAC
ATT
5′
'Template
strand'
DNA
Transcription
make RNA copy of upper DNA strand using lower
strand as a template (T residues correspond to U)
Ribosome-binding site:
influences amount of protein
produced from mRNA
'Start' codon
5′
AUG
UAA
3′
One of three
'stop' codons
mRNA
fMet is incorporated as the N-terminal
amino acid residue (this is often
removed later in the final protein)
Translation
make protein encoded by mRNA strand
N fMet
C
Protein

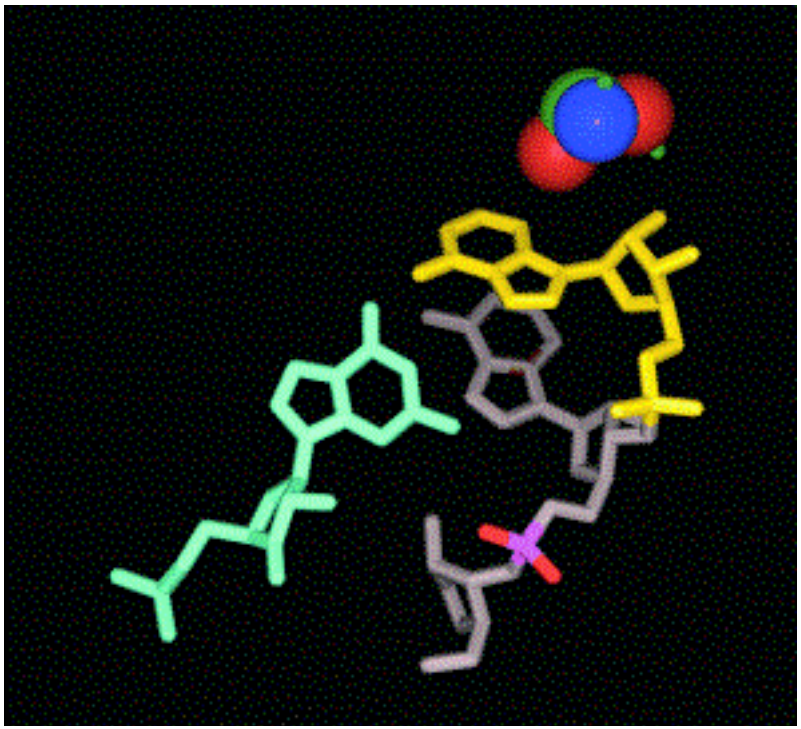
The genetic code Remember **U** in mRNA and **T** in DNA

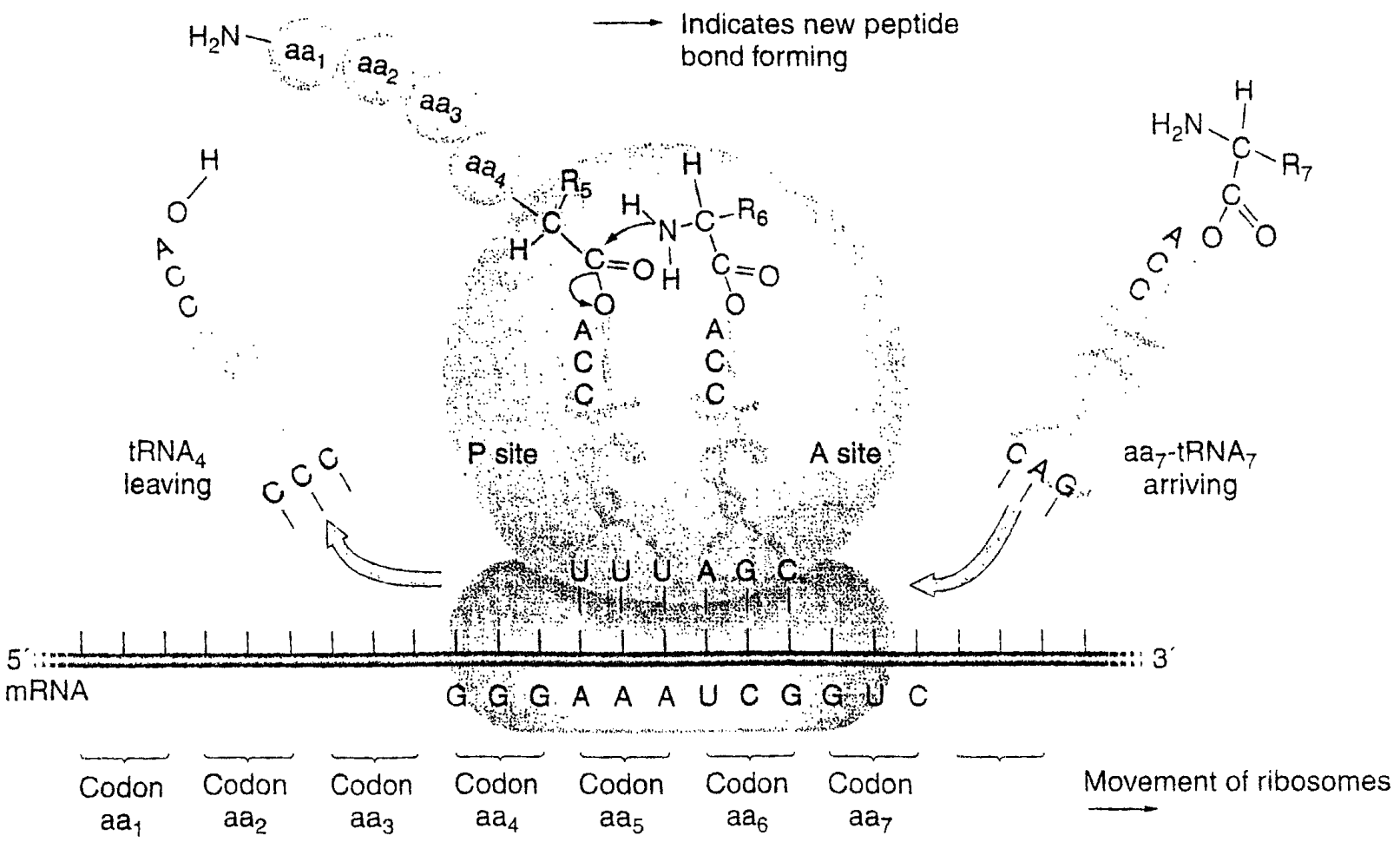
5′ base	Middle base				3′ base
	U	**C**	**A**	**G**	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA Leu	UCA Ser	UAA Stop*	UGA Stop*	**A**
	UUG Leu	UCG Ser	UAG Stop*	UGG Trp	**G**
C	CUU Leu	CCU Pro	CAU His	CGU Arg	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	**A**
	AUG Met†	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	**U**
	GUC Val	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA Glu	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

*Stop codons have no amino acids assigned to them.

†The AUG codon is the usual initiation codon as well as that for methionine residues elsewhere. The code is almost universal but differences have been found in mitochondrial DNA from some organisms


50S
pep
R1
R2
C
O
N
H
H2
mRNA
P
A
30S

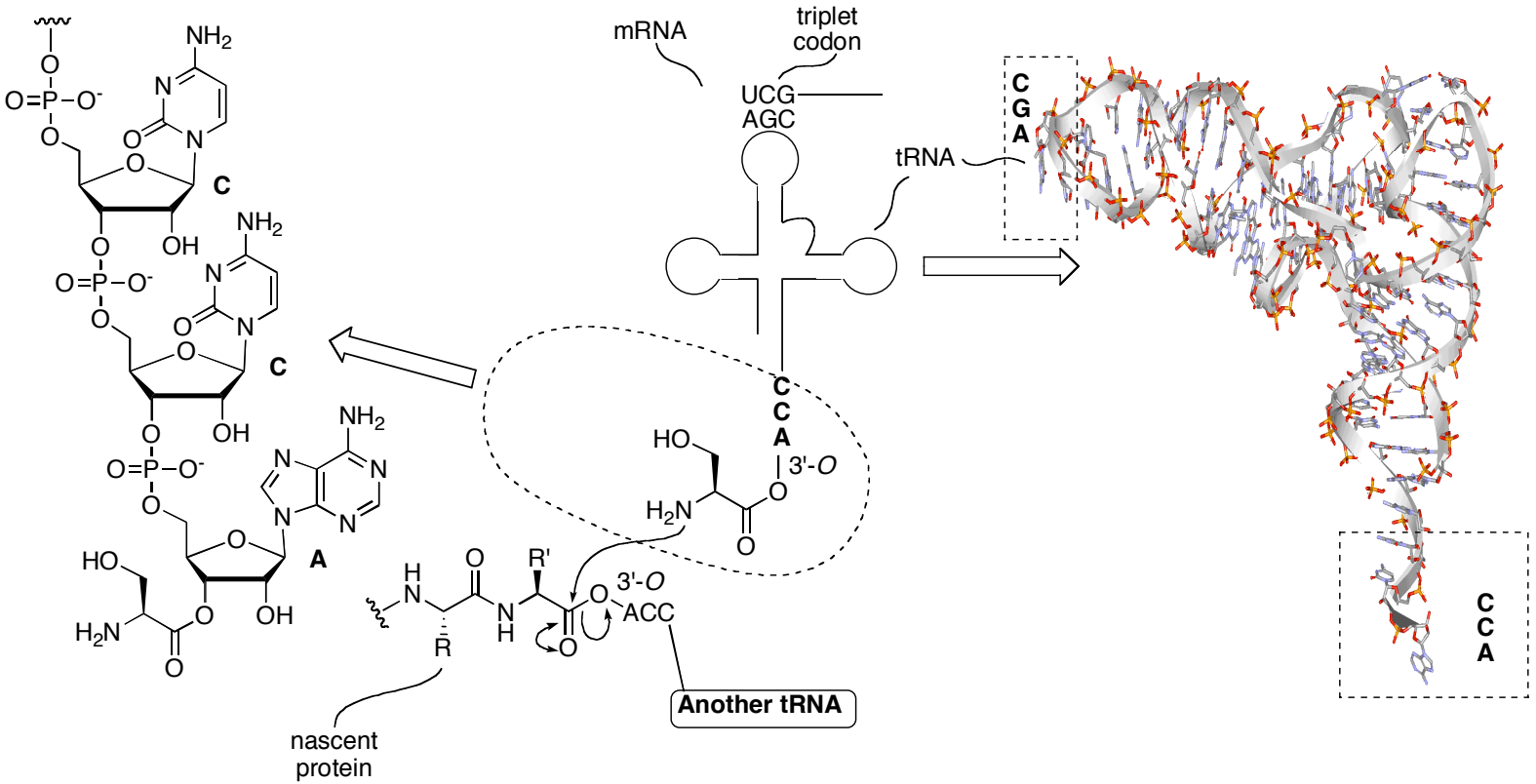


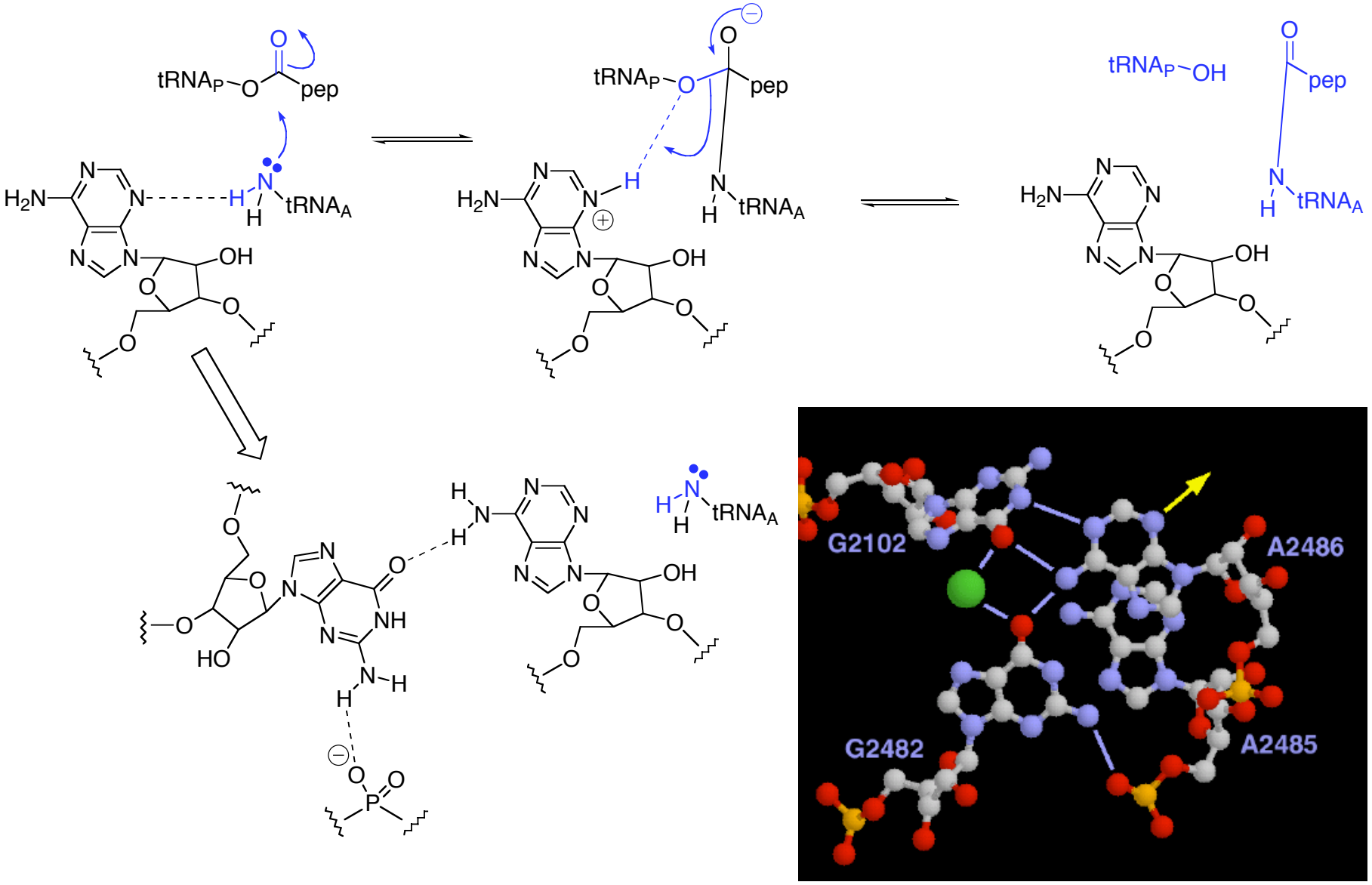


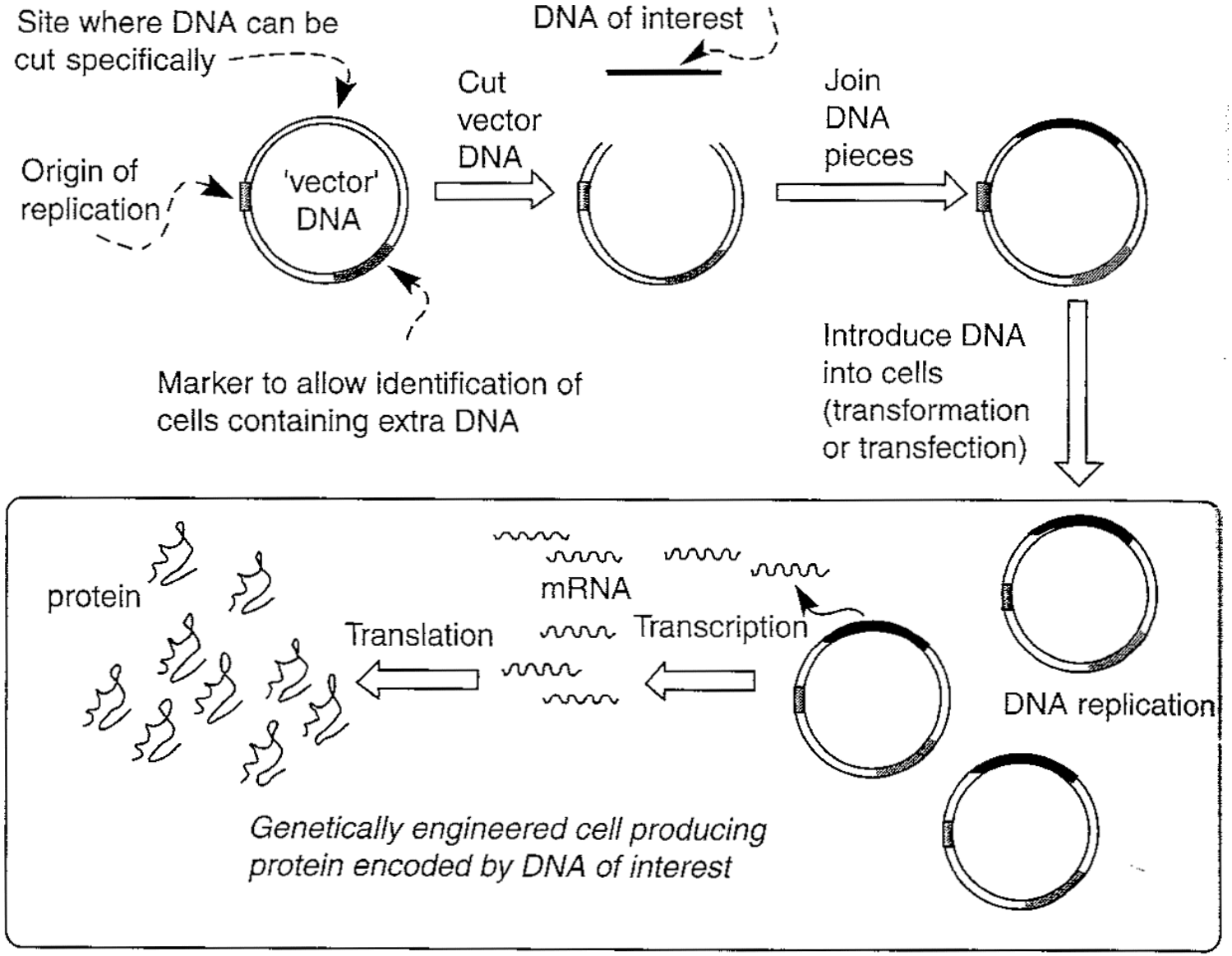
.....|UUU|UUA|GGG|AAA|UCG|GUC| .....
..... Phe Leu Gly Lys Ser Val

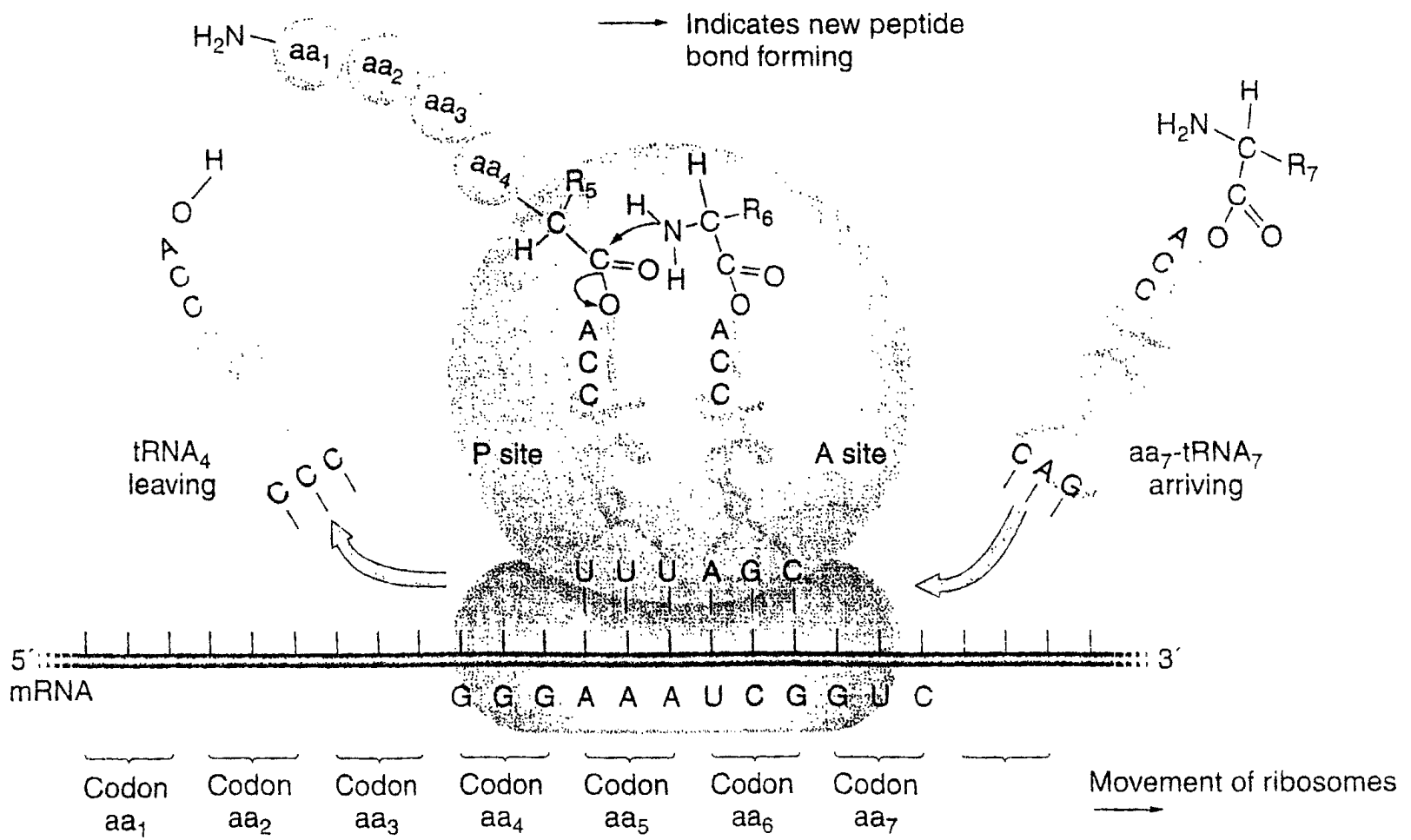
⇒

.....|UUU|UUA|GGG|AAA|U**GC**|GUC| .....
..... Phe Leu Gly Lys **Cys** Val


mRNA
triplet codon
UCG
AGC
tRNA
C
G
A
C
C
A
3'-O
HO
H2N
O
C
C
A
nascent protein
Another tRNA
3'-O
ACC
R
R'
H
N
O
NH2
OH
HO
H2N
A
C
C
C
A


tRNAP
O
pep
H2N
N
H
tRNAA
OH
HO
NH
P
tRNAP–OH
G2102
A2486
G2482
A2485

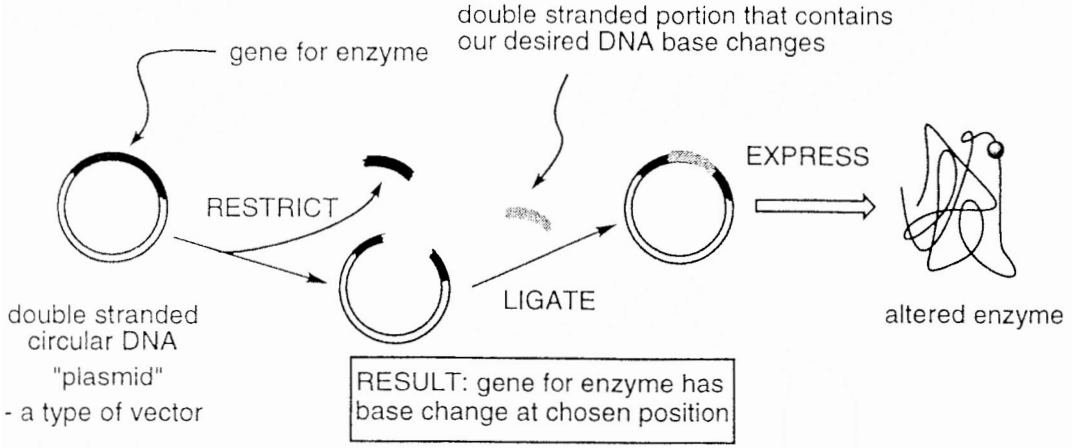

Site where DNA can be cut specifically
DNA of interest
Cut vector DNA
Join DNA pieces
Origin of replication
'vector' DNA
Marker to allow identification of cells containing extra DNA
Introduce DNA into cells (transformation or transfection)
protein
mRNA
Translation
Transcription
DNA replication
Genetically engineered cell producing protein encoded by DNA of interest



.....|UUU|UUA|GGG|AAA|UCG|GUC| .....
..... Phe Leu Gly Lys Ser Val

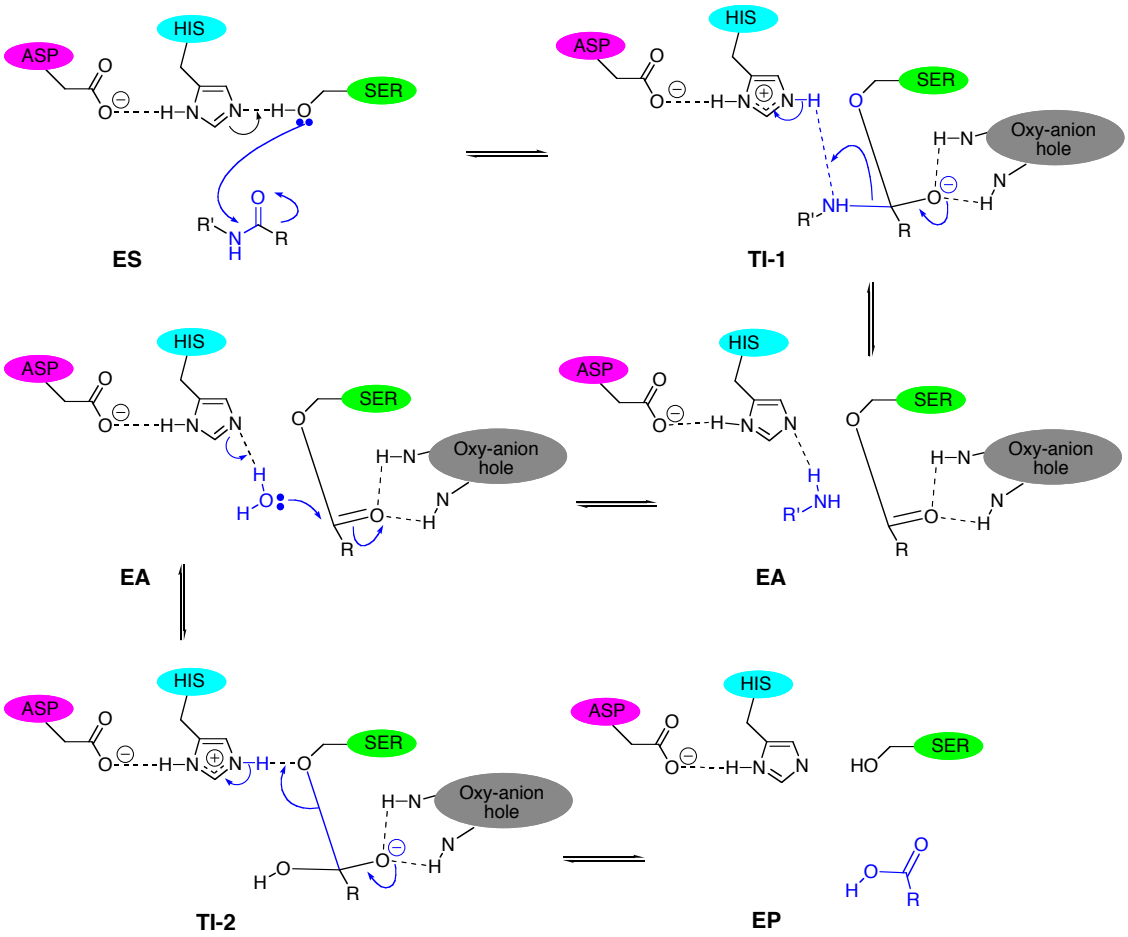
.....|UUU|UUA|GGG|AAA|U**GC**|GUC| .....
..... Phe Leu Gly Lys **Cys** Val

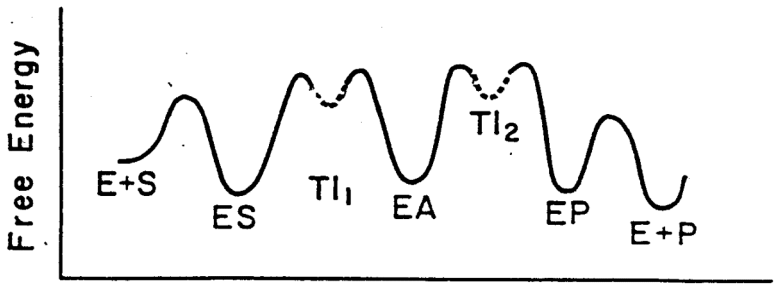
Site-Directed Mutagenesis



First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

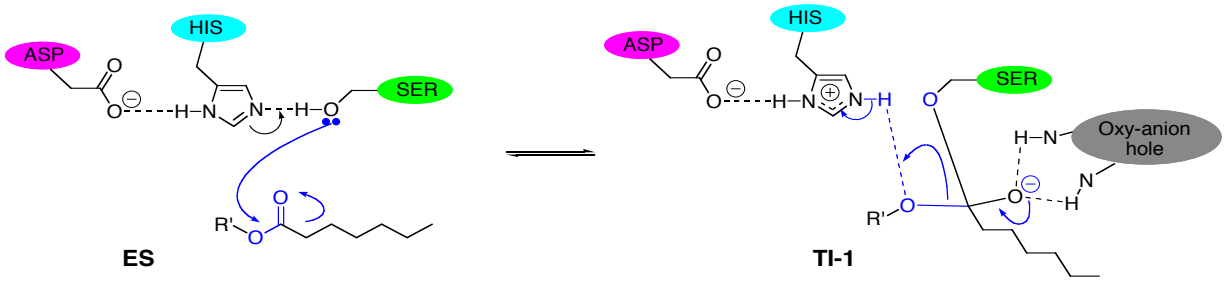
Serine Hydrolases





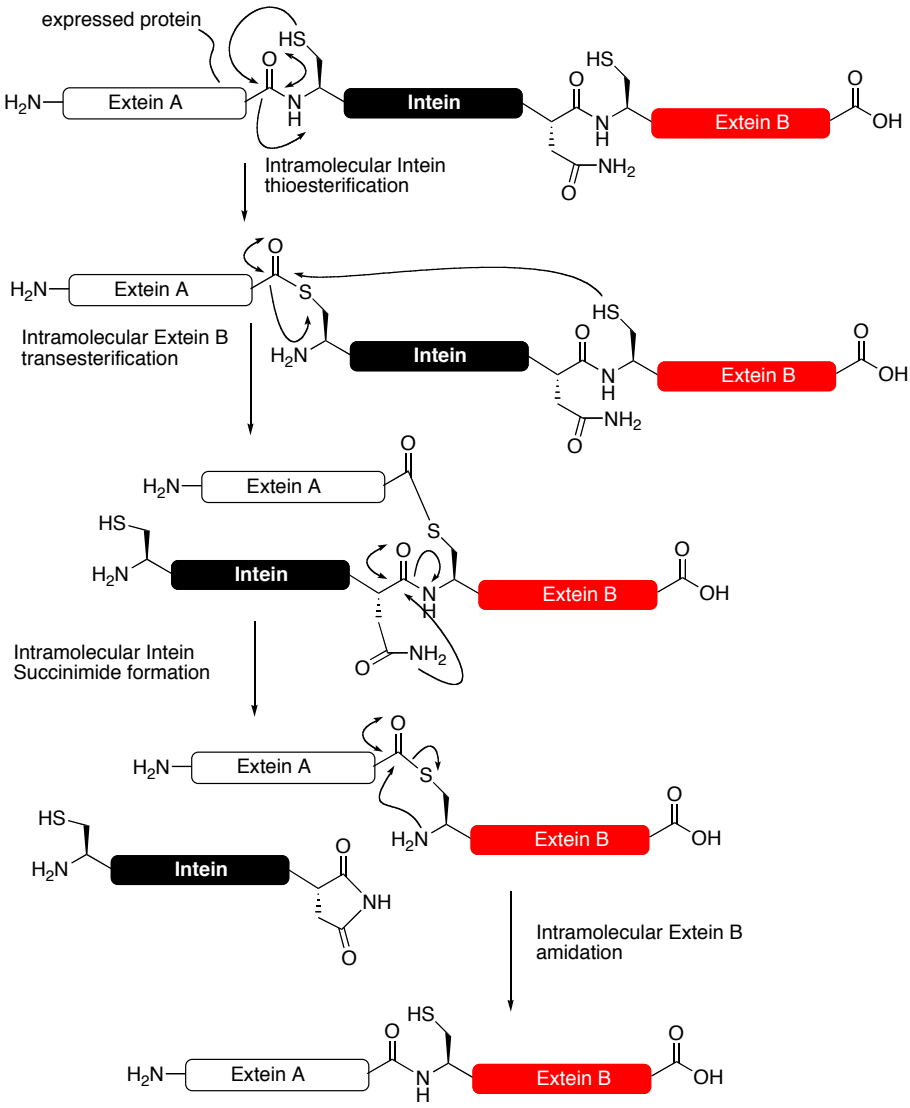
A representation of the expected free energy diagram for serine proteinase catalysis. From evolutionary principles the free energies of all the transition states are expected to be similar, and the energies of all the intermediates are anticipated to be similar

• The amide bond could equally be an ester bond and the same mechanism is used by many lipases:



Protein Splicing - Inteins and Exteins

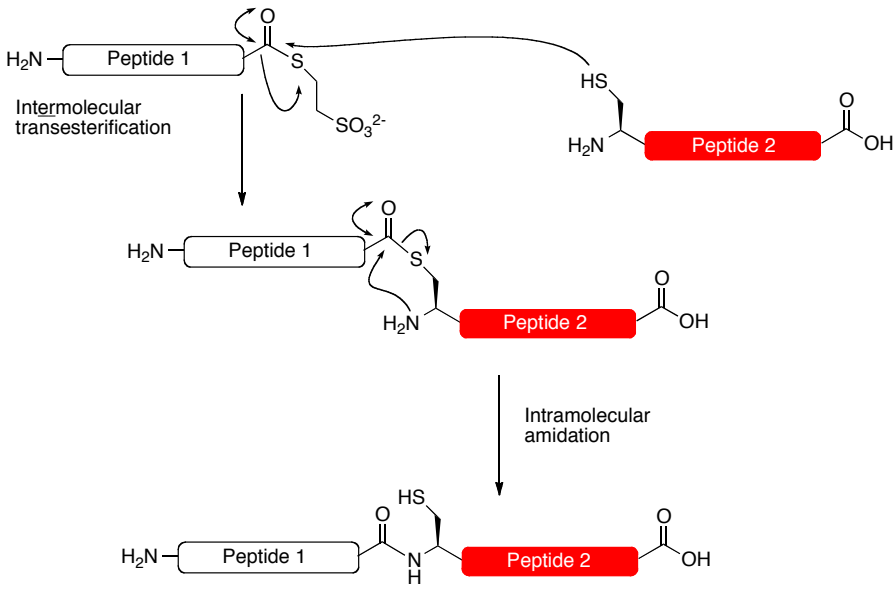
• The same principles apply in natural, post-translational protein splicing mechanisms.
• These autocatalytic processes involve nucleophilic catalysis and proximity effects.
• They lead to the 'chopping out' of a peptide segment called an *intein* and result in the joining of *exteins*. The system is in some ways analogous to *exons* and *introns* in pre-mRNA RNA.
• In this example, an Asn amide side chain acts as a nucleophile to cleave the intein by forming a succinimide. Examples with Gln are also known.



'Native Chemical Ligation'

• This can form the basis for a method for a *chemical* method of making proteins with a 'native' amide bond from peptide fragments: one a thioester (Peptide 1: how would you make this?) and the other a peptide with an *N*-terminal Cys residue (Peptide 2).

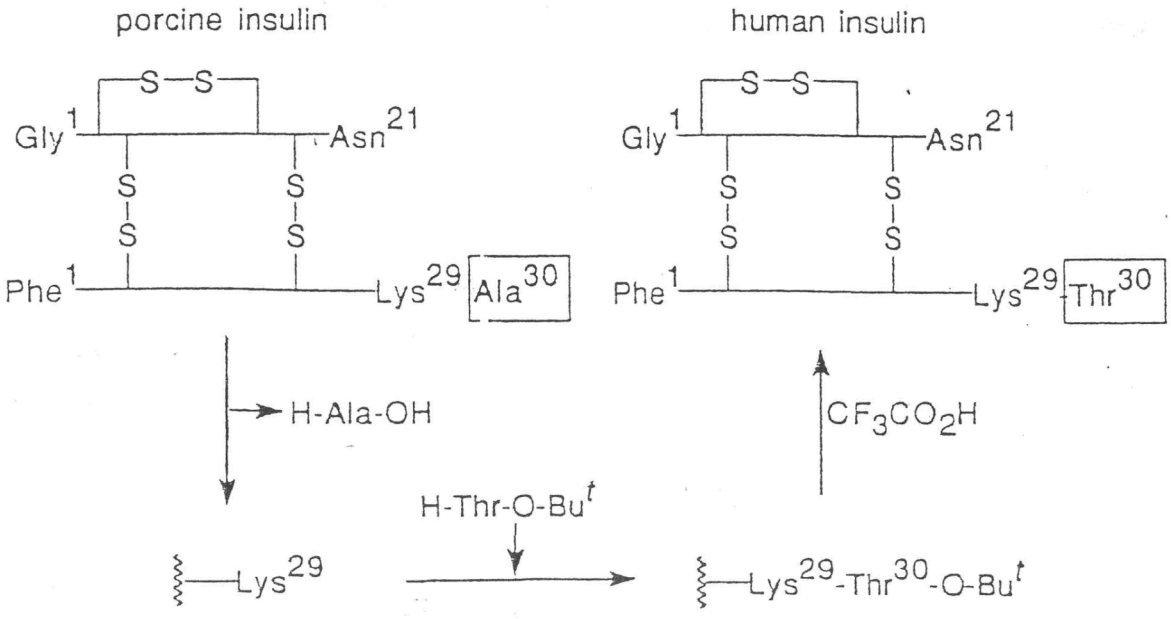
• These fragments could be made using standard methods and then assembled into larger structures, even proteins.



Common proteases and their preferred cleavage sites.

Protease	Type	Preferred Cleavage Sites
α-chymotrypsin and subtilisins	Ser	-Trp(Tyr,Phe,Leu,Met) ↓ Xaa-
elastase	Ser	-Ala(Ser,Met,Phe) ↓ Xaa-
pepsin	Asp	-Phe(Tyr,Leu) ↓ Leu(Phe)-
thermolysin	metallo	-Phe(Gly,Asp,Leu) ↓ Leu(Phe)-
papain	Cys	-Phe(Leu,Val)-Xaa ↓ Xaa-
trypsin	Ser	-Arg(Lys) ↓ Xaa-
clostripain	Cys	-Arg ↓ Xaa-
endoprotease Lys-C (*Achromobacter*)	Ser	-Lys ↓ Xaa-
endoprotease Glu-C (V8 protease)	Ser	-Glu (Asp) ↓ Xaa-
carboxypeptidase Y	Ser	-Xaa ↓ Xaa-OH
carboxypeptidase B	metallo	-Xaa ↓ [Arg,Lys]-OH
carboxypeptidase A	metallo	-Xaa ↓ [Asp,Glu,Phe,Leu]-OH
aminopeptidase M	metallo	H_2N-Xaa ↓ Xaa-
pyroglutamate-aminopeptidase	Cys	pGlu ↓ Xaa-
cathepsin C	Cys	H_2N-Xaa-Xaa ↓ Xaa-
proline iminopeptidase	Ser	Pro ↓ Xaa-

Enzymatic conversion of porcine into human insulin.



Digest followed by Mass Spectrometry

• Protein sample are often cleaved beforehand (typically by enzyme but could use other chemical methods) and MS is a complementary method in many cases

• Once a protein has been digested by a specific protease (e.g. trypsin) then these chunks can be examined by further fragmented by MS to get detailed sequence data

• Two main ionization methods: Electrospray Ionization (ESI) and Matrix-Assisted Laser Desorption Ionization (MALDI)

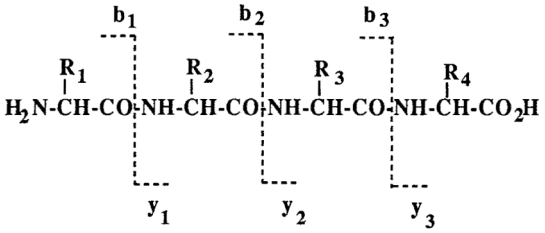
• MALDI is popular with crude samples e.g. cell fractions or lysates but ESI often gives better precision

• Ions can be selected on the basis of m/z (mass/charge) and then diverted into a separate chamber or second MS. Thus this technique is known as MS-MS. The first MS to measure all ions, the second to measure the 'mass selected ions'

• Sequencing can be achieved by Collision Induced Dissociation (CID) of peptide linkages.

• This can be achieved by varying the potential on the ions formed to accelerate them with good control when exposed to a collision gas introduced into the spectrometer.

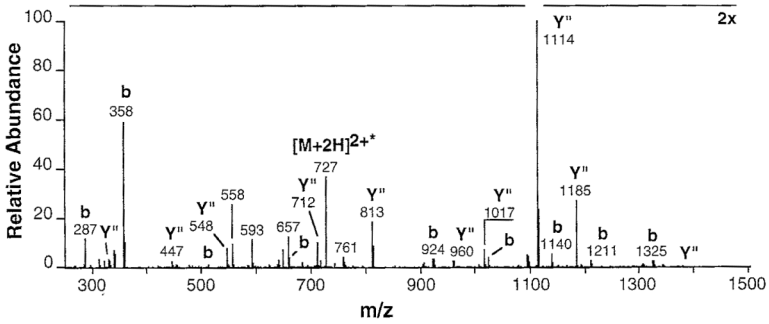
• Cleavage can result in an acylium (C=O+) – so-called Y series ions – or ammonium (NH+) – so called B-series ions.



Examples

b_1 ion **y_3 ion**

• Databases now exist of common fragmentations linked with known sequences such that most modern sequencing spectrometers will not only sequence but also identify your protein from genome information.



T G Q A P G F T Y D A N K

n	b	Y"	n
0	1.0	**1470.7**	14
1	102.1	**1369.6**	13
2	159.1	**1312.6**	12
3	**287.1**	**1184.6**	11
4	**358.2**	**1113.5**	10
5	**455.2**	**1016.5**	9
6	**512.2**	**959.4**	8
7	**659.3**	**812.4**	7
8	**760.4**	**711.3**	6
9	**923.5**	**548.3**	5
10	**1024.5**	**447.2**	4
11	**1139.5**	**333.2**	3
12	**1210.5**	261.2	2
13	**1324.6**	147.1	1
14	1452.7	19.0	0

This fragment came from an EndoLys cleavage.
The MH_2^{2+} peak (m/z 727) was 'mass selected'