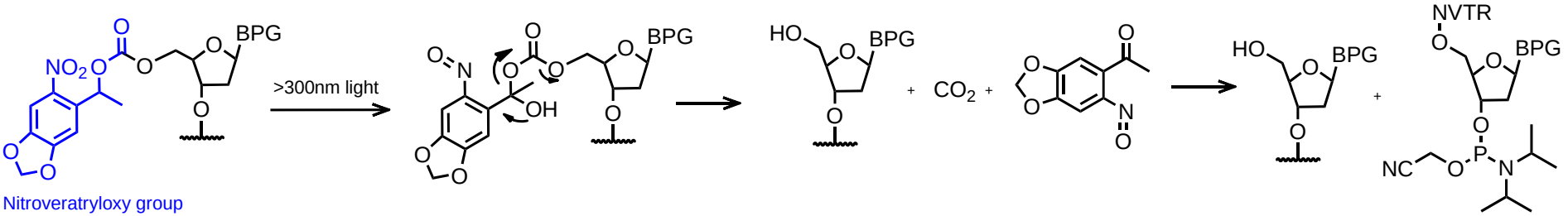
Lecture 5- DNA Arrays and Introduction to Protein Structure and Function

Light directed synthesis of oligonucleotides

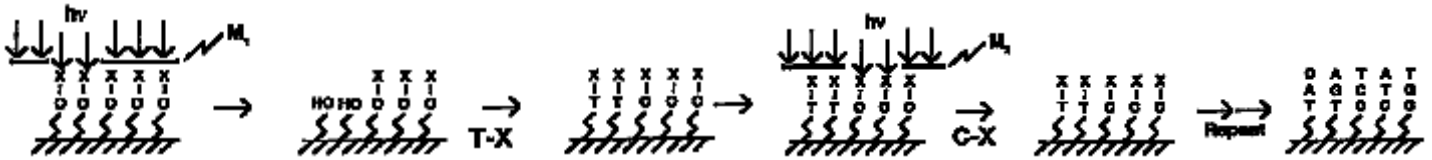


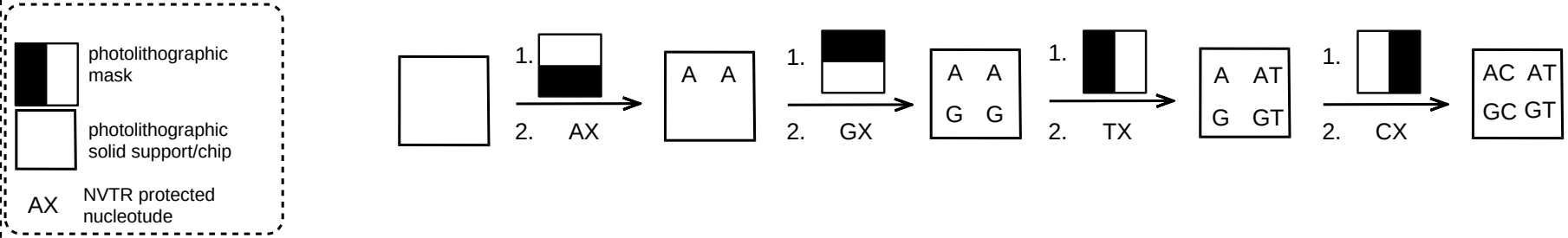
Nitroveratryloxy group

Can now undergo another round of coupling with deprotected OH and another NVTR protected nucleotide

Spatial control of oligonucleotides synthesis using masking

- Masking specific regions of the photolithographic chip during light exposure allows for spatially controlled deprotection and coupling.

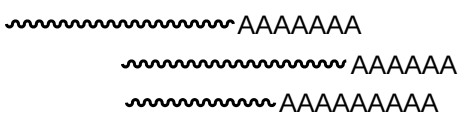




spatial resolution ~1-10um per site. Therefore a 1x1cm chip has on the order of 10^4-10^6 distinct sites for probe hybridization.

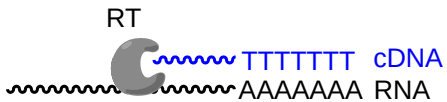
Example-using microarrays to profile breast cancer cells

1 - Isolate mRNA from breast cancer cells



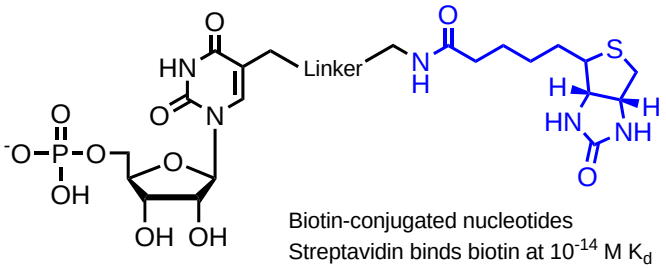
-mRNAs are poly adenylated
-can be isolated using polyT column

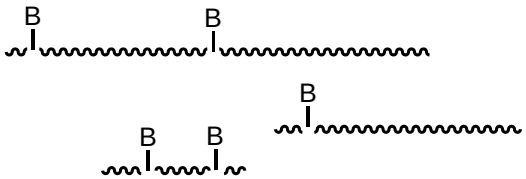
2- Reverse transcription using reverse transcriptase to generate cDNA



Reverse transcriptase catalyzes completementary DNA strand formation (cDNA)

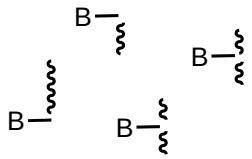
3- In vitro transcription with biotin-conjugated nucleotides





Biotin-tagged in vitro transcription products

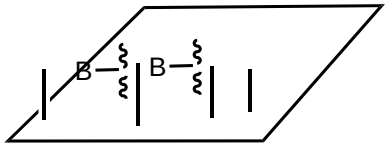
4-Fragment biotin-containing transcription products



Fragmented biotin-tagged in vitro transcription products

5- Hybridize fragmented products on DNA array

6- Wash and detect using fluorophore-conjugated streptavidin



-Every gene has multiple detection probes within the chip to assure accurate representation of mRNA expression
-Can detect expression of transcripts at the genome-wide level (~30k mRNA, miRNAs, etc)

In general, allows for discovery of over-expressed proteins and can inform therapeutic design. For example, if HER2 is over-expressed in the breast cancer sample, one can generate drugs like Herceptin ADCs.

Proteins

Central dogma of biology

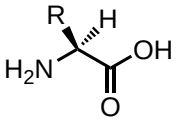
Proteins have 20 building blocks (amino acids) and can form more complex structures

DNA → mRNA → Protein

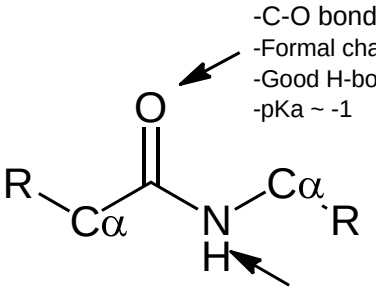
DNA and mRNA only have 4 building blocks each

Amino acids

-20 canonical amino acids
-All amino acids are translated in L configuration, although epimerization to D configuration can occur in certain circumstances



Amide linkage properties



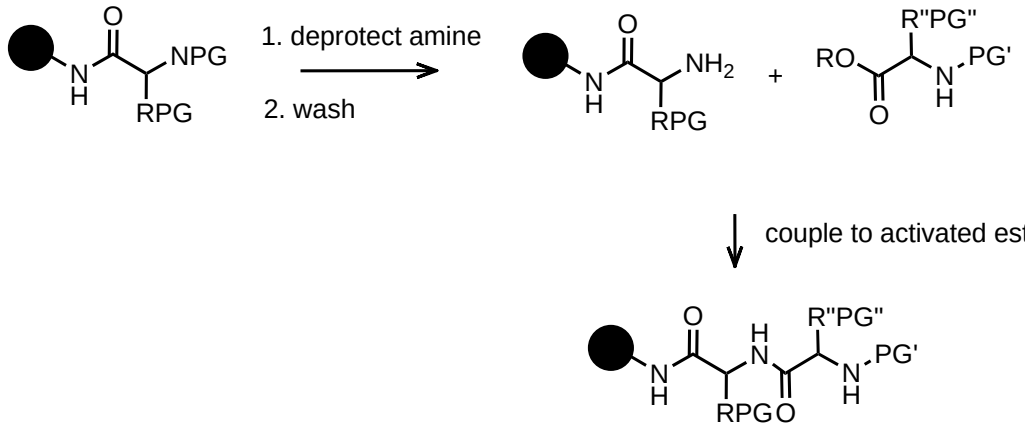
-C-O bond length 1.23 Å
-Formal charge -0.42
-Good H-bond acceptor
-pKa ~ -1

-N-H bond length 1.33 Å
-Formal charge +0.2
-Good H-bond donor, relatively acidic
-pKa ~ 15-20

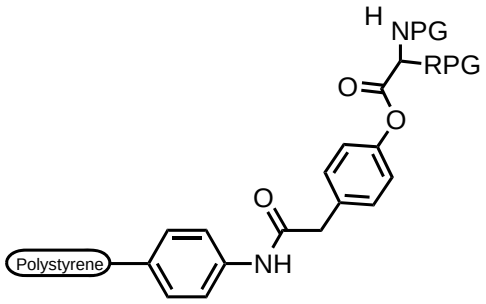
Solid phase synthesis of peptides

General scheme

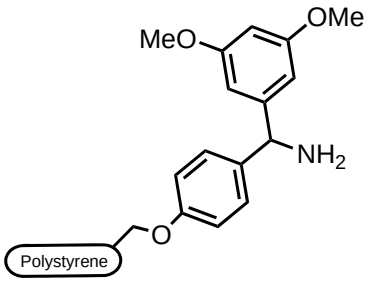
-all synthesized on solid support, allows for rapid purification
-linkages to resin are either amide or ester
-peptides are synthesized C to N, opposite of translation
-protecting groups for side chains and amines must have orthogonal deprotection strageties



Resins



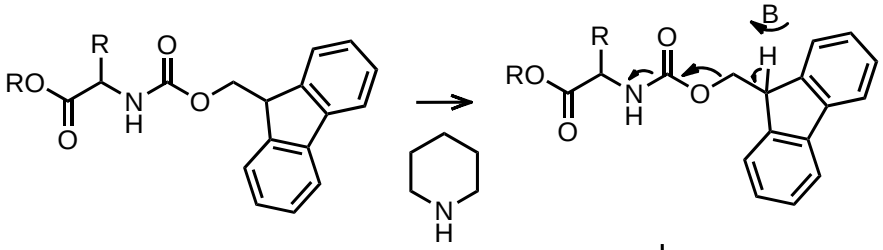
PAM Resin
-peptide starts with ester linkage
-acid labile
-when cleaved leaves carboxylate C-terminus



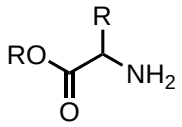
RINK Resin
-peptide starts with amide linkage
-acid labile
-when cleaved leaves amide C-terminus

Amino group protection

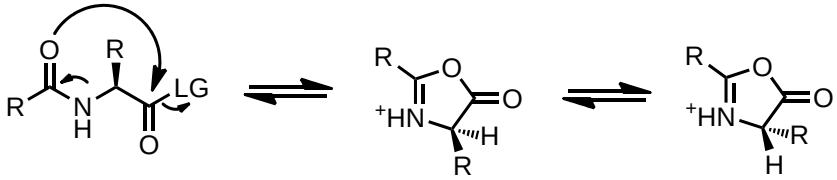
FMOC
-Fluorenylmethyloxycarbonyl
-Cleaved with mild base



-Piperidine used as base (generates aromatic carbanion).
-Stronger bases are not used because they can induce epimerization of Cα by abstracting the alpha proton



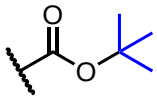
Carbamates are frequently used instead of amides due to potential cyclization products



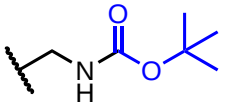
Carbamtes lead to less acidic alpha C-H iminium ion product and therefore minimize racemization

Side chain protecting groups
-All are acid labile

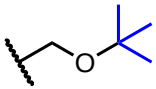
Carboxylates-
Asp, Glu



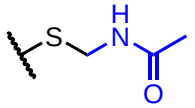
Lysine-



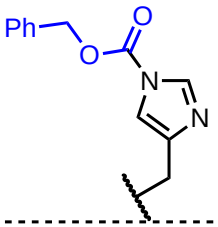
Alcohols-
Ser, Thr



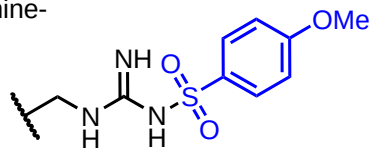
Cysteine-



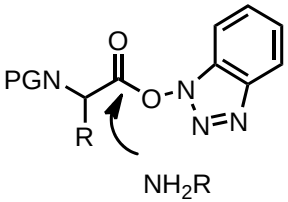
Histidine-



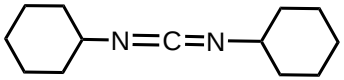
Arginine-

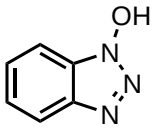


Carbodimide coupling strategy

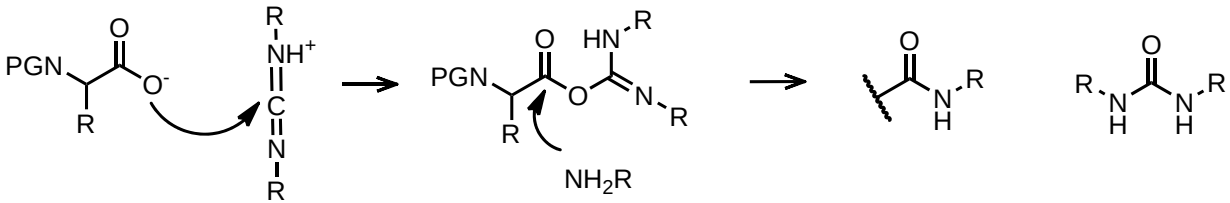


DCC





Hydroxybenzotriazole is used as an additional exchange reagent to activate ester

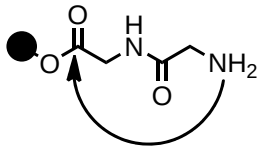


-needs to be free acid
in order to protonate DCC

Urea formed is an insoluble white precipitate which is easily purified

Other nuances in peptide synthesis

- Sometimes one must consider the sequence of the peptide when choosing resin for synthesis. e.g. Gly,Gly as the first two residues in the peptide may result in cyclization



-Another concern is the aggregation of resin bound peptide intermediates

Deprotection of side chains and resin with acid

Deprotection mixture includes TFA, Ethanediothiol (used to scavenge tert-butyl cations), and Anisole.

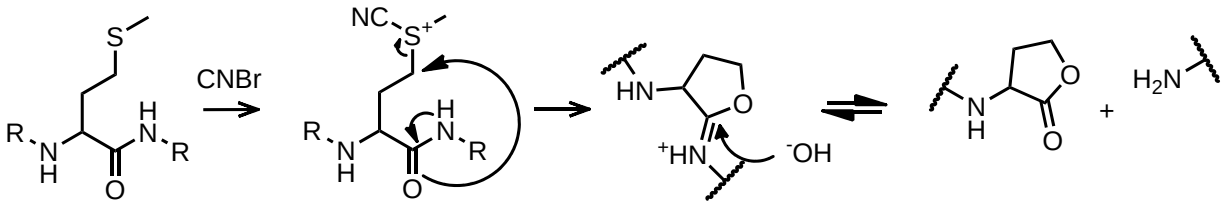
Protein sequencing

Methods for sequencing proteins include:
1. Mass spectrometry
2. Chemical degradation

Either method requires fragmentation which can be performed:
1. Enzymatically - e.g. trypsin cleaves specifically at Lys and Arg
2. Chemically - e.g. CNBr degradation

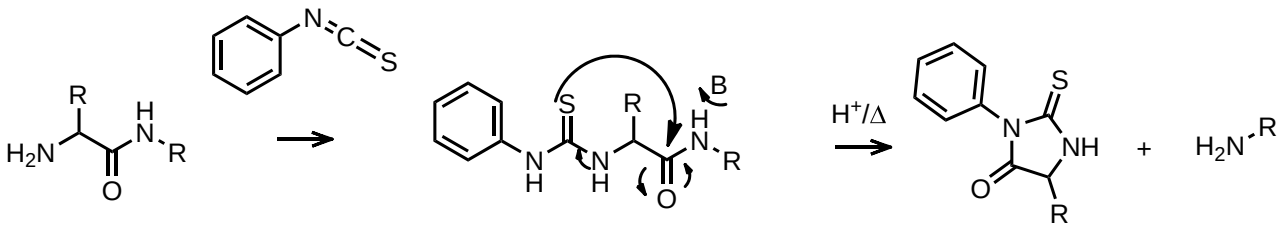
CNBr degradation

-cleaves at methionine residues



Edman degradation for protein sequencing

-Cleaves first N-terminal residue only



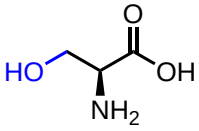
-Each phenylthiohydantoin-conjugated amino acid can be identified via chromatographic methods

-Leaves an intact N terminus

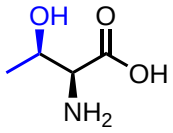
Side chains of amino acids

Alcohols

pKa ~ 16



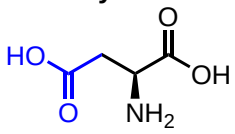
Serine



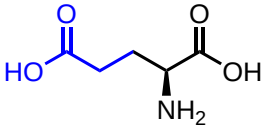
Threonine

-Ser/Thr need activation by general base to form alkoxide species to act as nucleophiles

Carboxylates



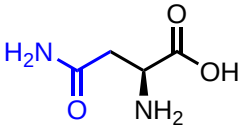
Aspartic Acid



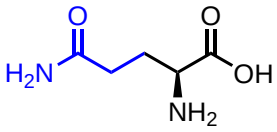
Glutamic Acid

-pKa 4-5
-Act as general acids or bases
-Important in formation of salt bridges

Amides



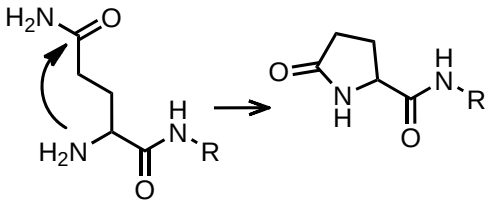
Asparagine



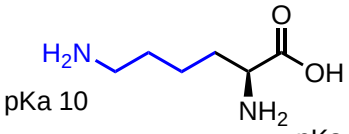
Glutamine

-Important for H-bonding
-Infrequent roles in catalysis

- glutamine can cyclize to form pyroglutamine in proteins if on N terminus



Lysine



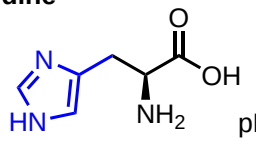
pKa 10

pKa 7-8

- can get selective modification of N terminal alpha amino if K is first residue due to lower pKa of alpha amino group

- can modify lysine residues with NHS-esters, used in ADCs

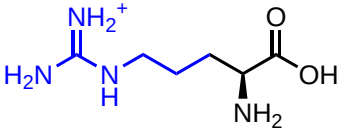
Histidine



pKa-6-7

-Can serve as a base to aid in catalysis
-Can be modified with chloroacetamides

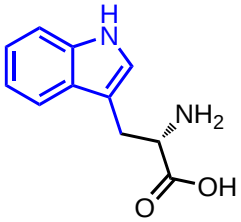
Arginine



pKa ~12

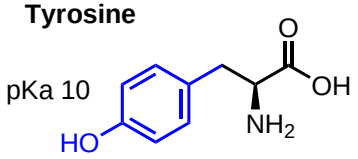
-Most frequently utilized in molecular recognition via its frequent use in salt bridge formation

Tryptophan

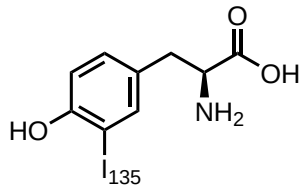


-acts an intrinsic spectroscopic probe for proteins
-Abs at 280nm λmax
-emission at 348nm λmax

Tyrosine

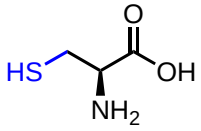


pKa 10



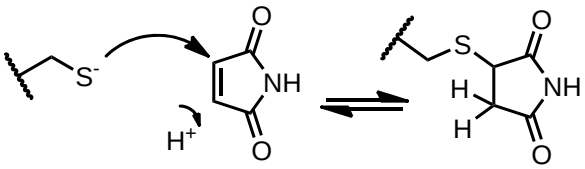
-good H-bond donor
-can serve as a nucelophile via formation of phenoxide anion (e.g phenoxide formation in topoisomerase mechanism)
-can be used to label proteins using I^{135} (beta emission)

Cysteine



pKa ~12

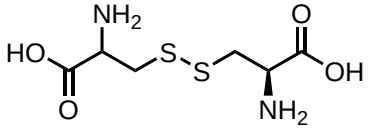
-best nucleophile among canonical 20 aa
-can selectively modify cysteine with maleimides



-maleimide conjugation can be used for ADC synthesis but is slowly reversible due to acidic nature of maleimide proton and the leaving group ability the thiolate

-can also be modified with chloroacetamides

- cysteines are oxidized to form cystines, which can act as inter and intra strand subunit disulfides to stabilize proteins



Cystine

-the 90 degree bent conformation is highly preferred