

CSE 427

Computational Bio

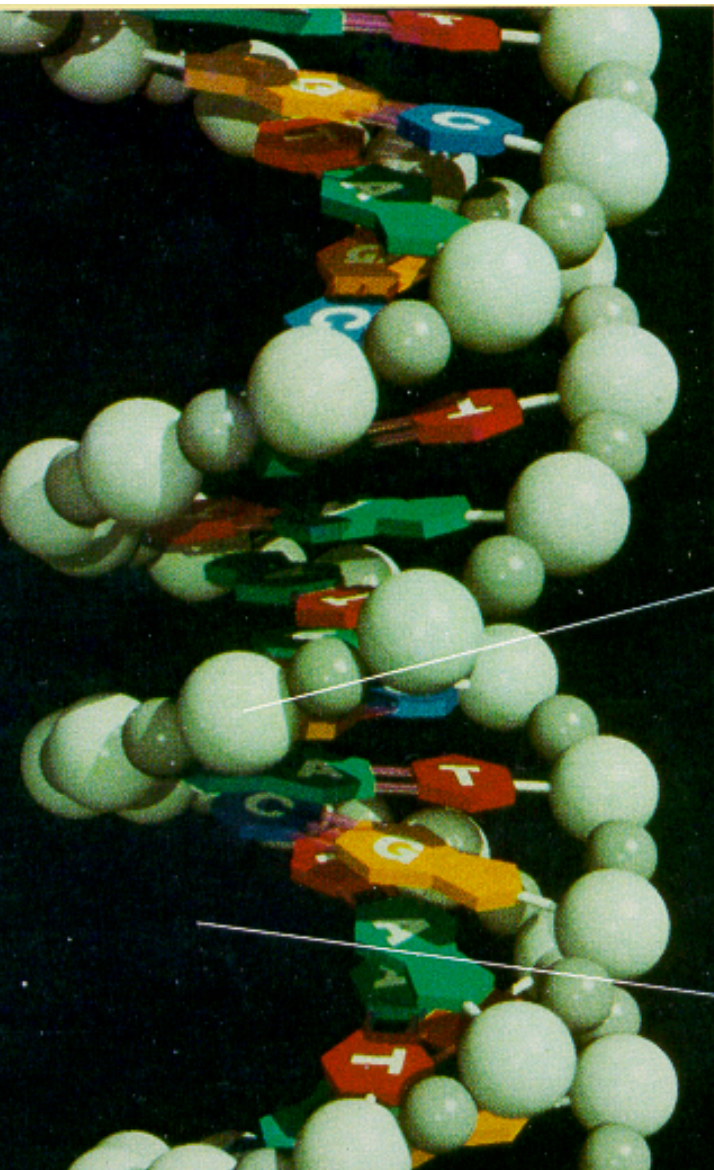
Gene Regulation

L. Ruzzo

DNA Binding Prote

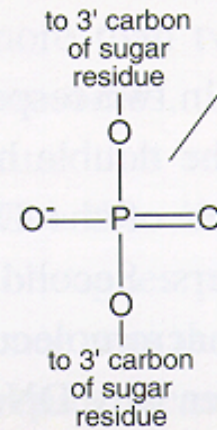
A variety of DNA binding proteins (“transcription factors”; a significant perhaps 5-10%, of all human proteins) modulate transcription of protein coding genes

The Double Helix

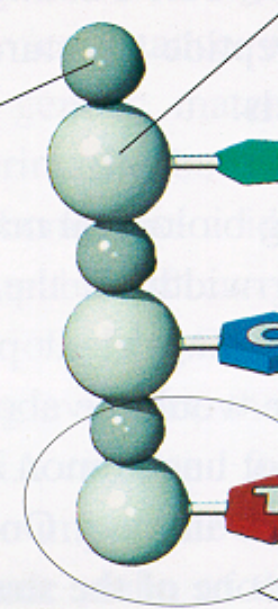


(a) Computer-generated Image of DNA (by Mel Prueitt)

(b) Uncoiled DNA Fragment



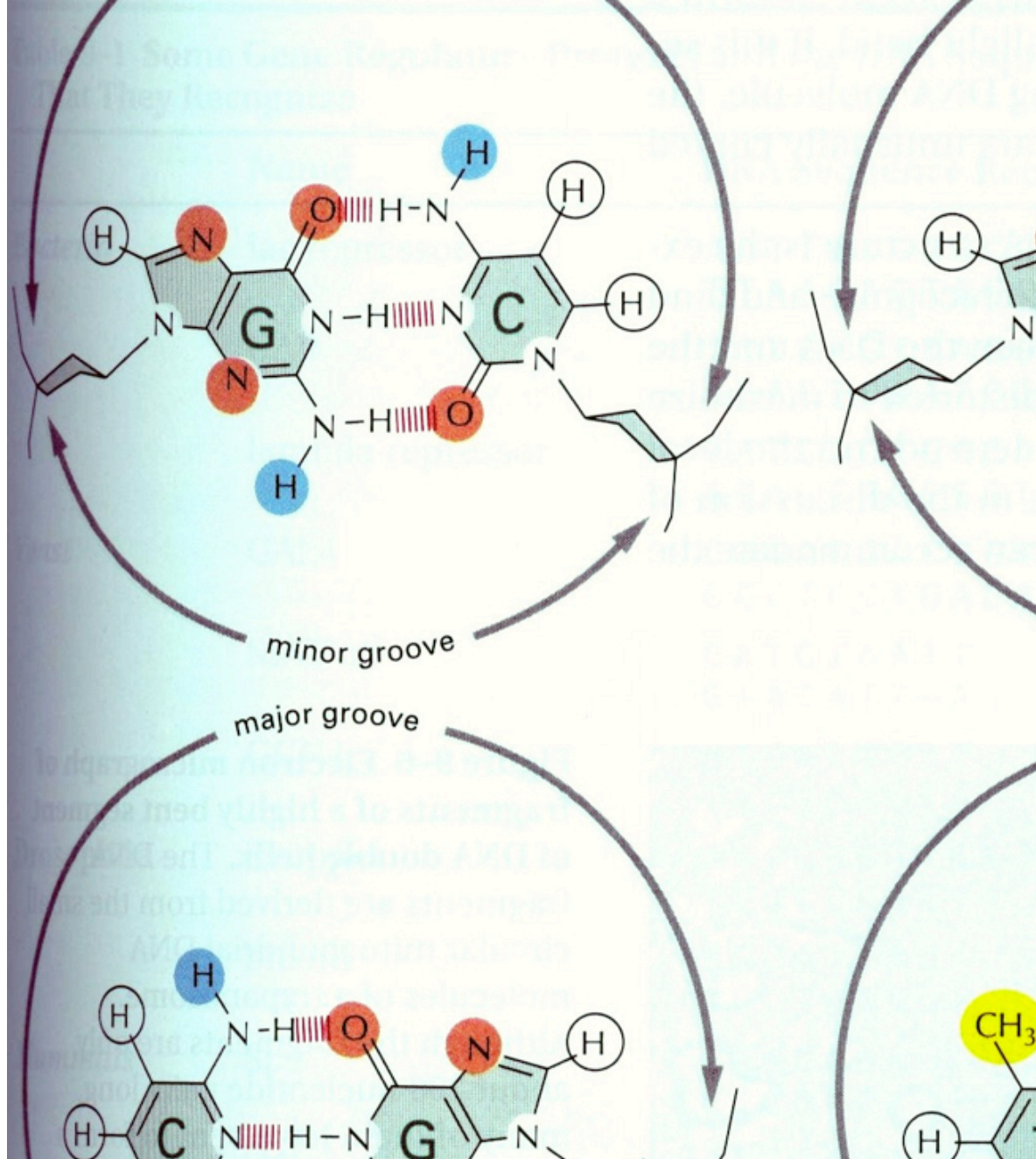
Phosphate group



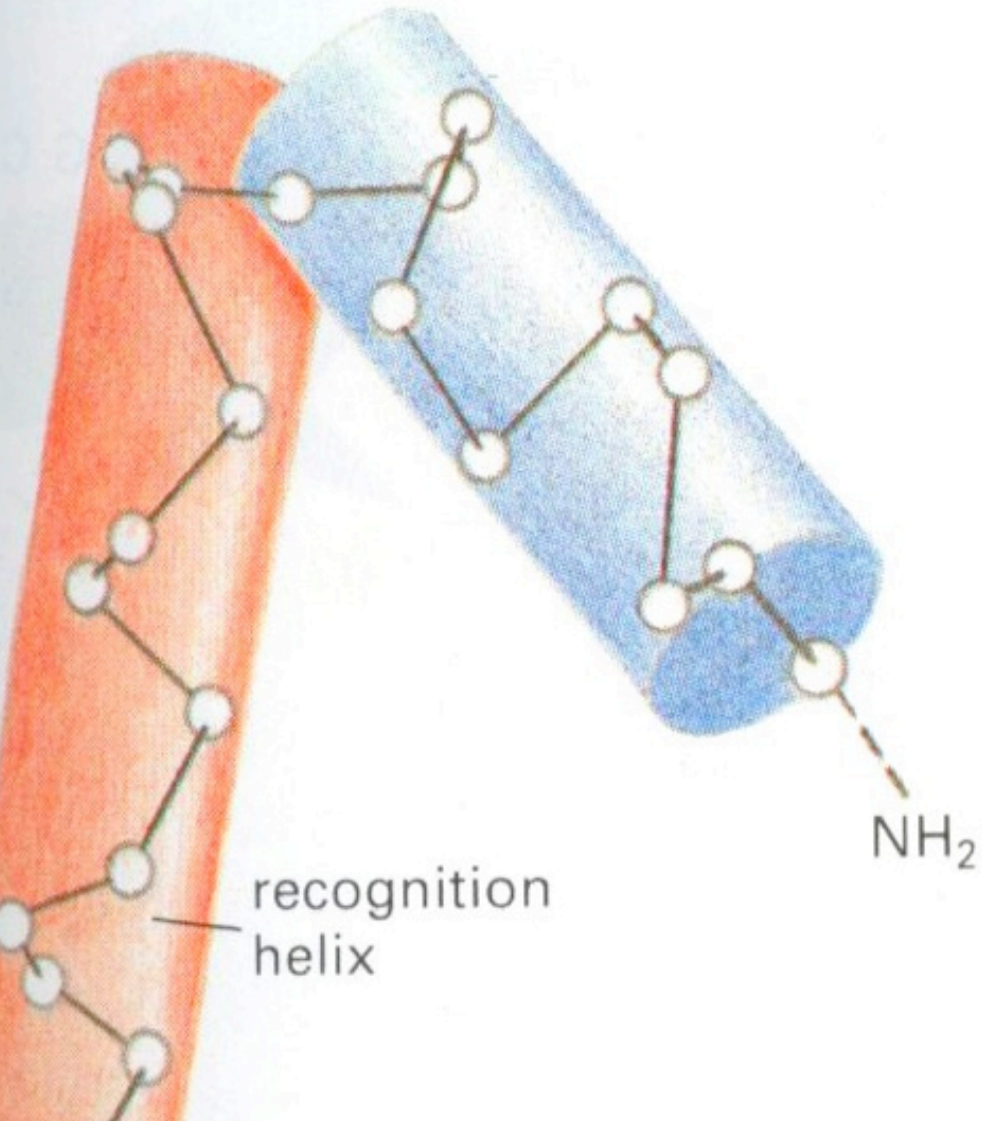
As shown, the two strands coil about each other in a fashion such that all

the
groove

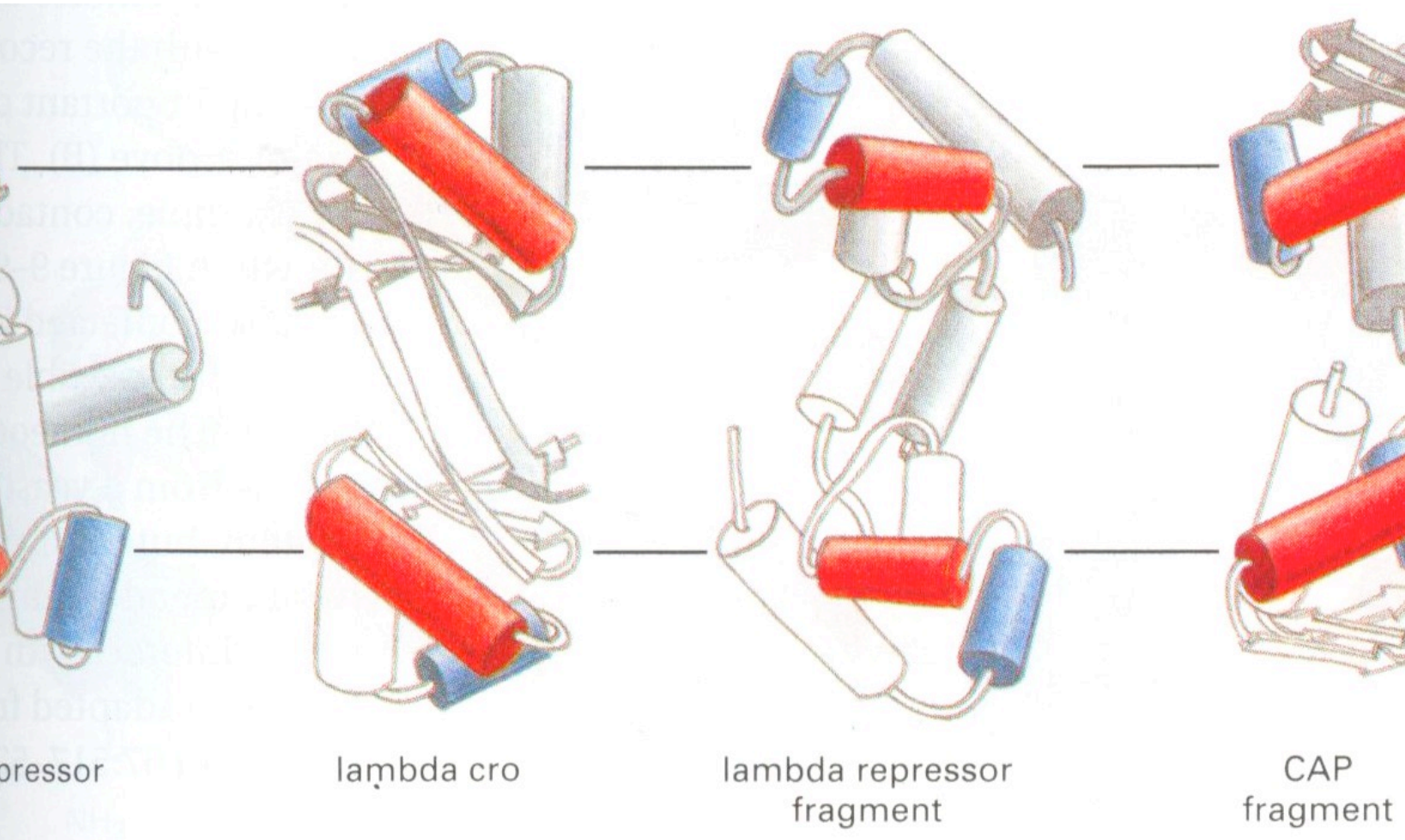
different
patterns of
potential H
bonds at
edges of
different
base pairs,
accessible

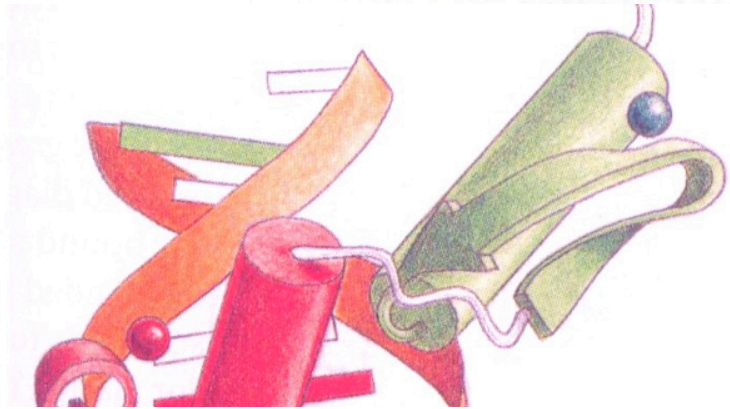
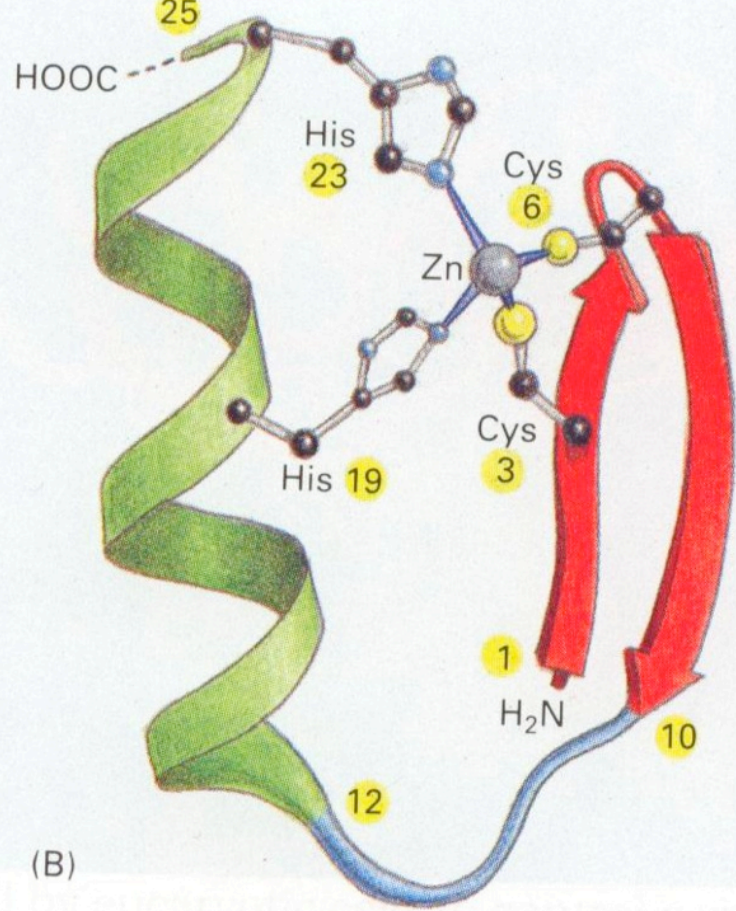
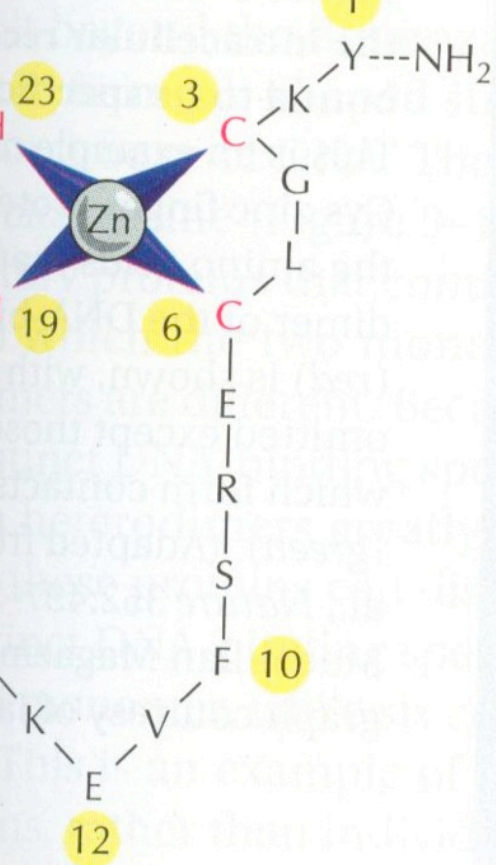


Helix-Turn-Helix DNA Binding



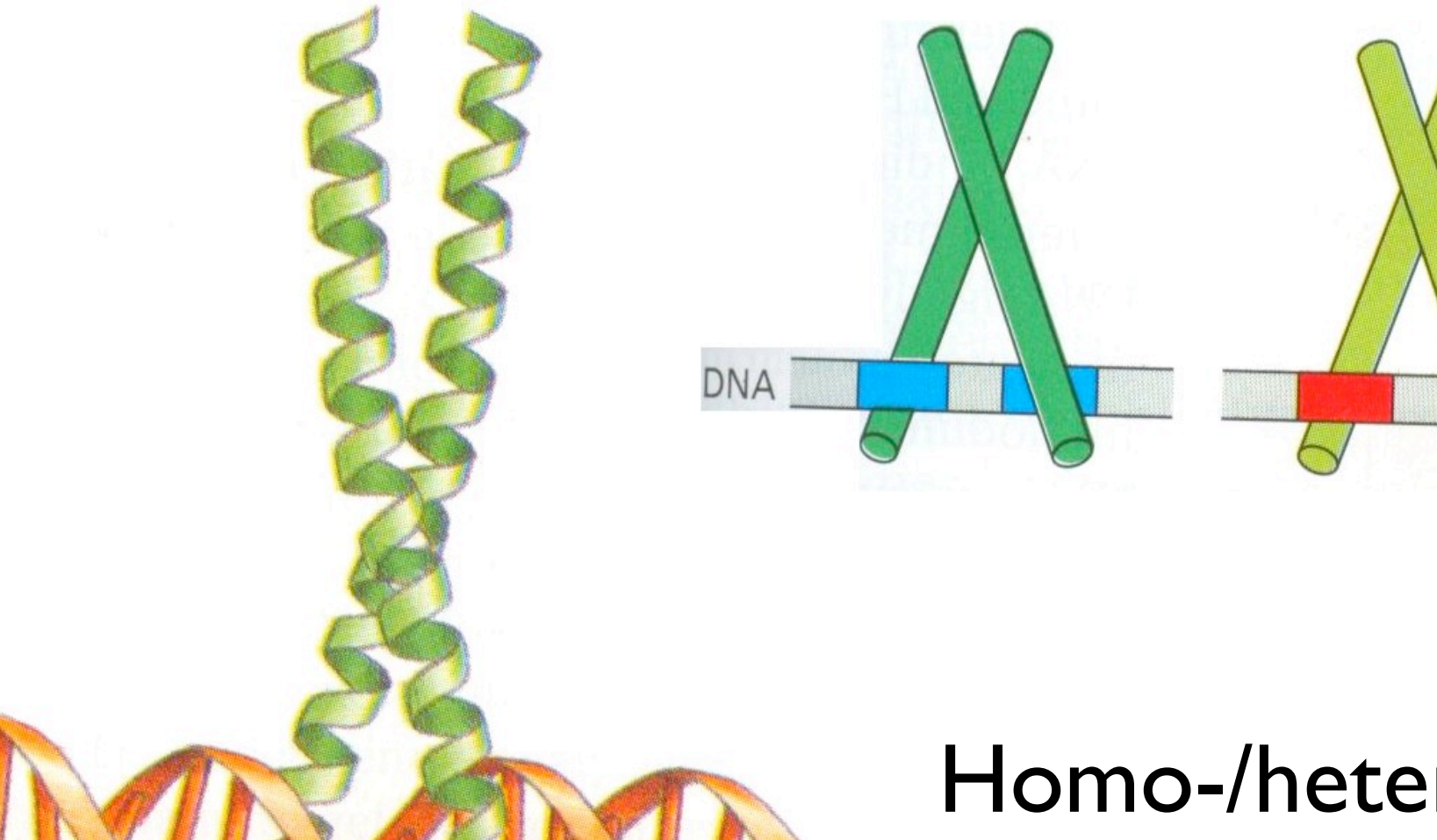
H-T-H Dimer



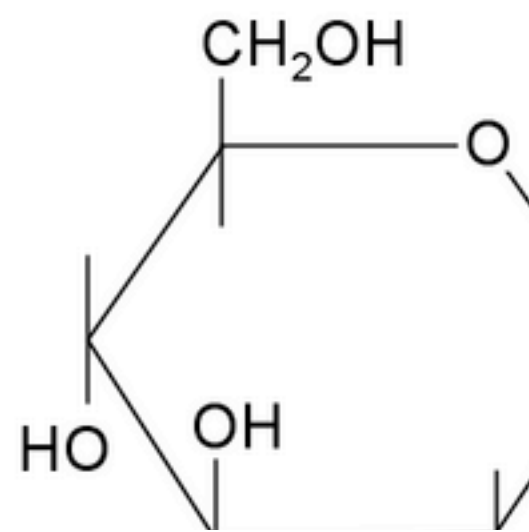
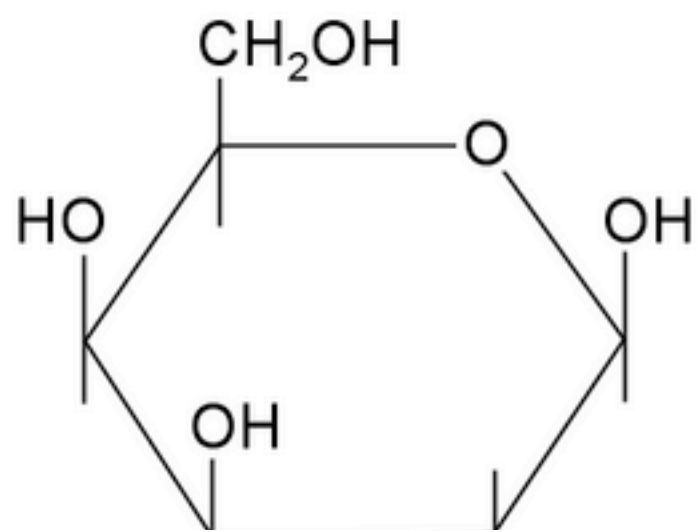
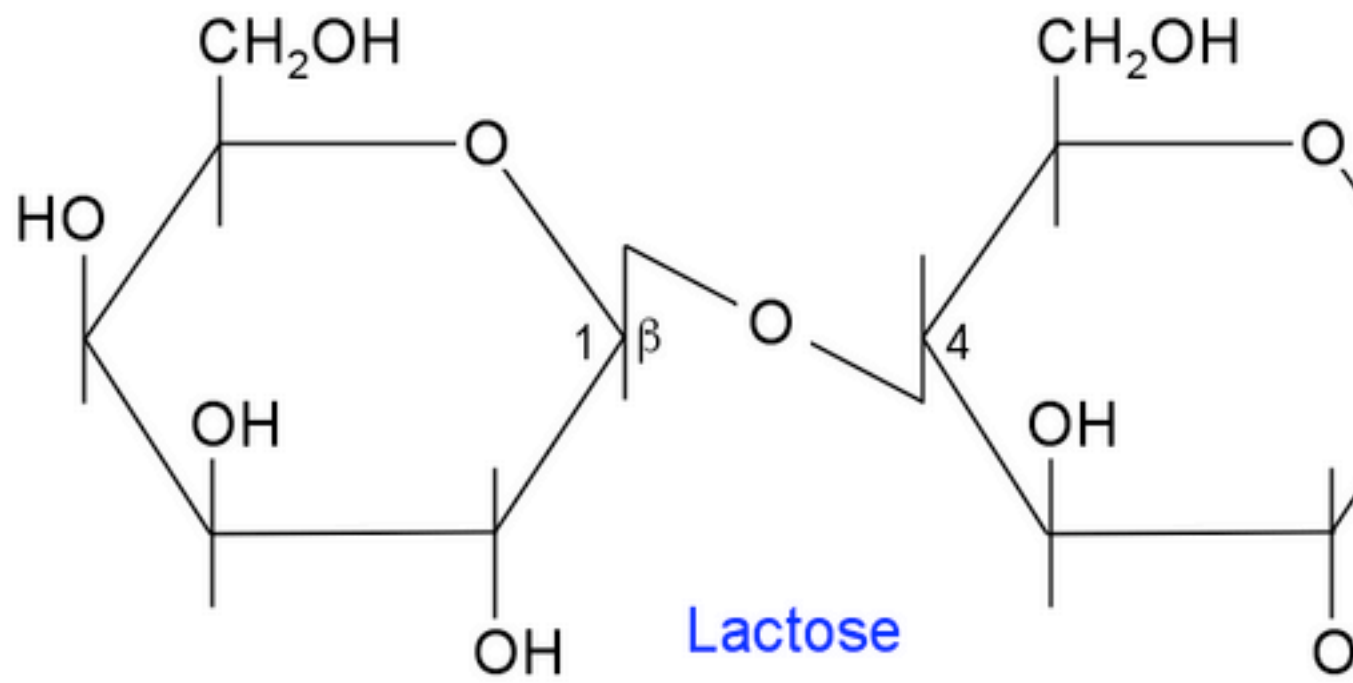


F

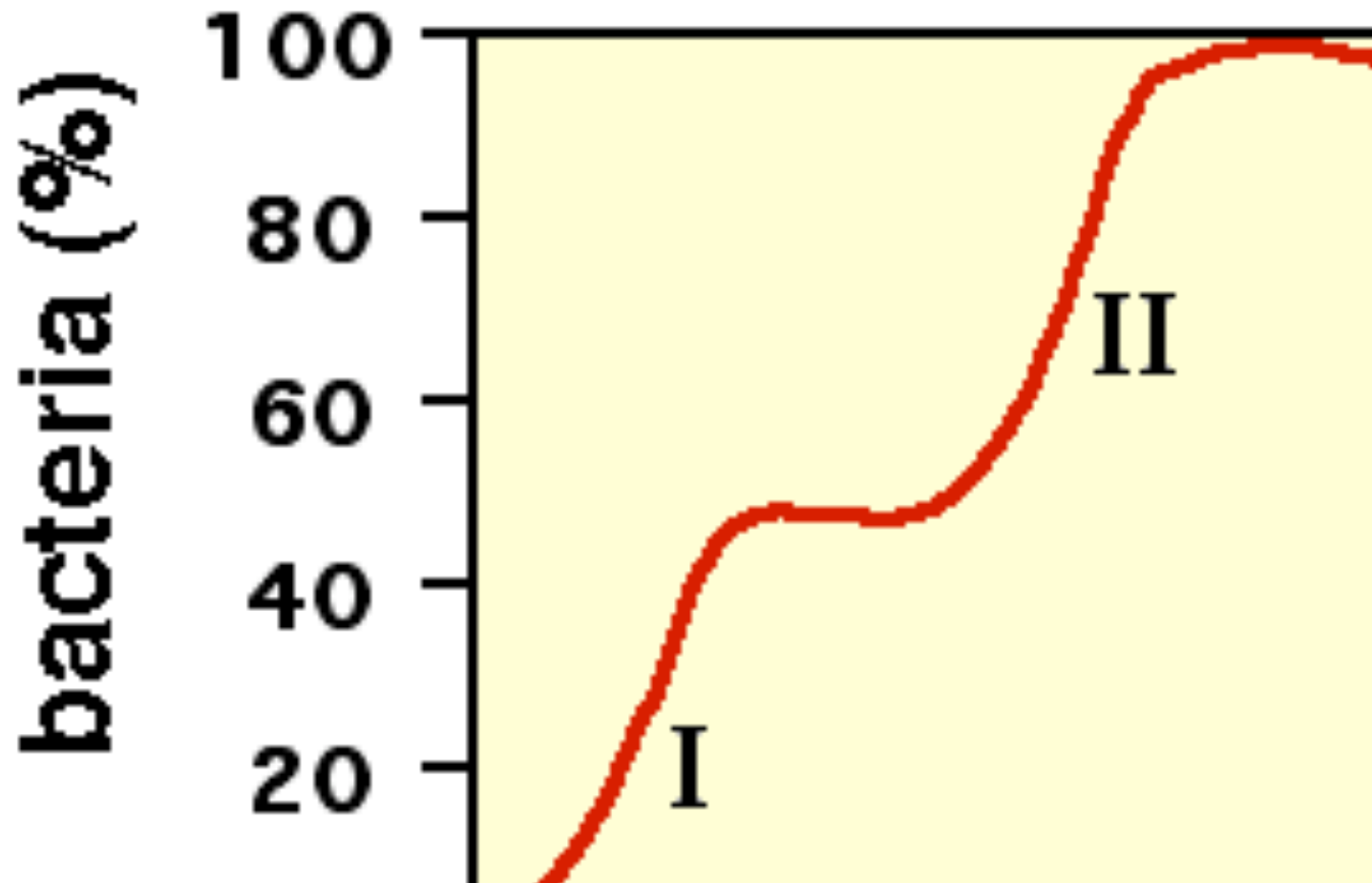
Leucine Zipper M

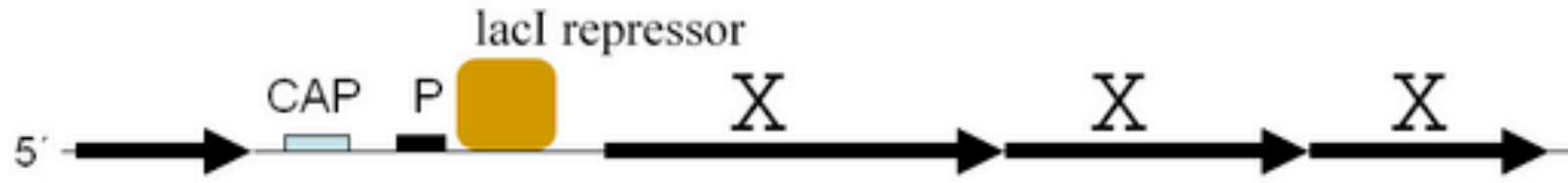
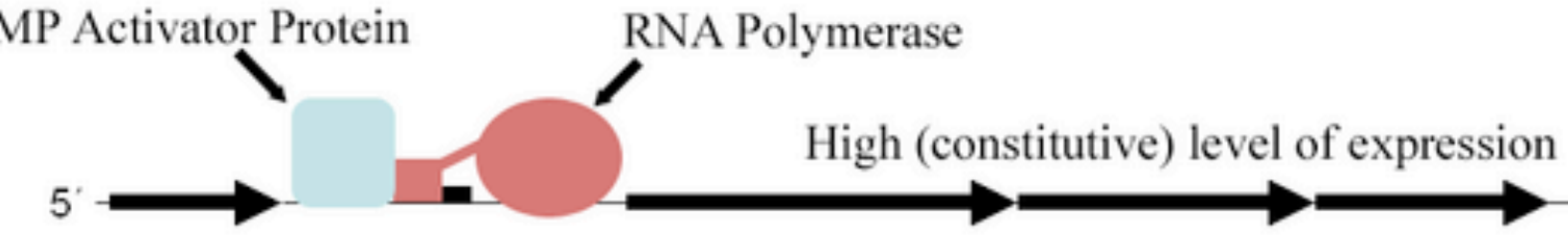
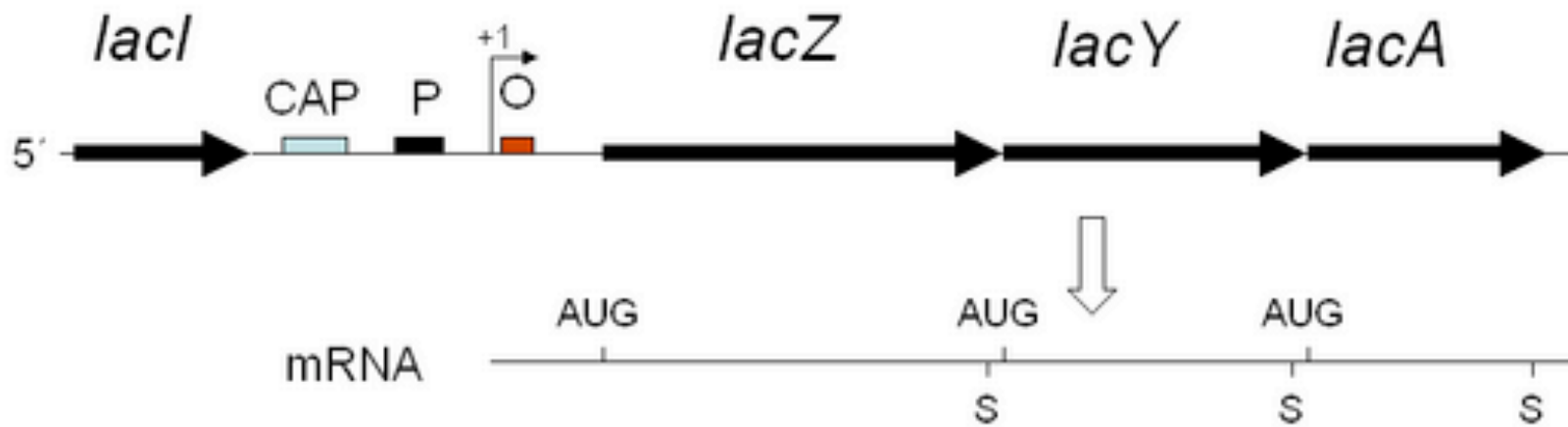


Homo-/heter

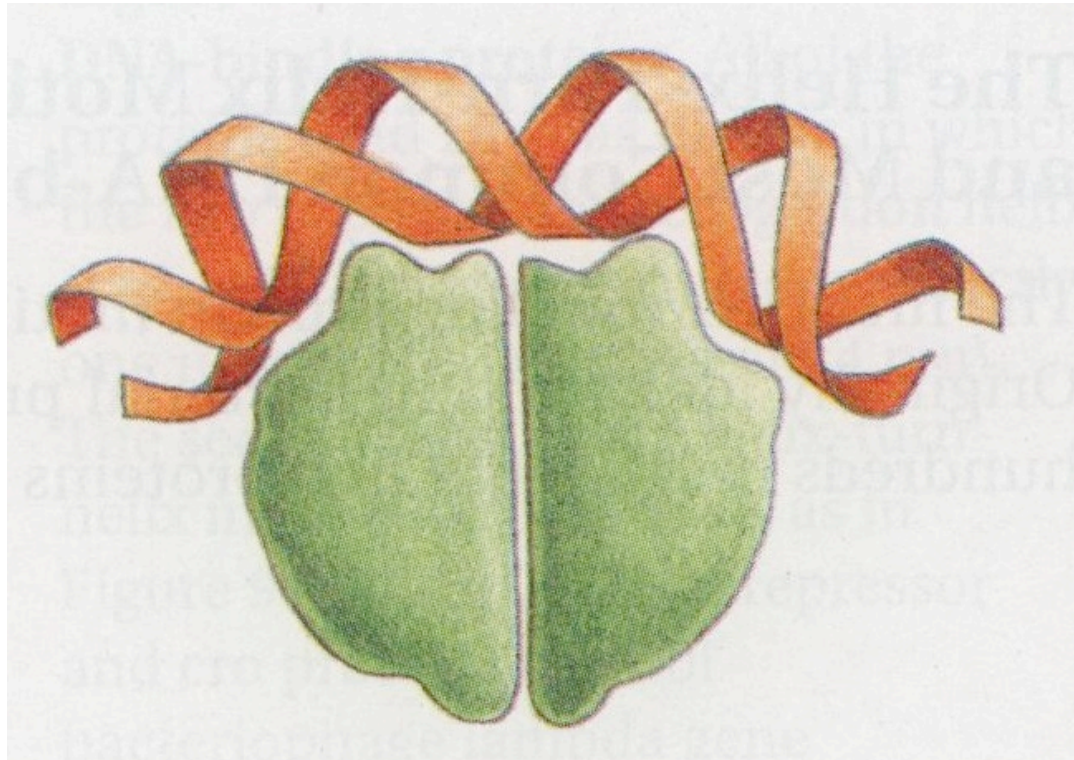


E. coli growth: glucose + lactose



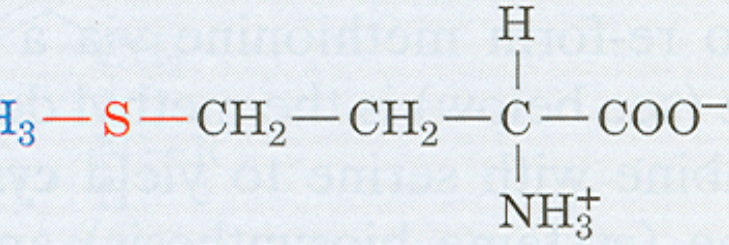


CAP bends DN

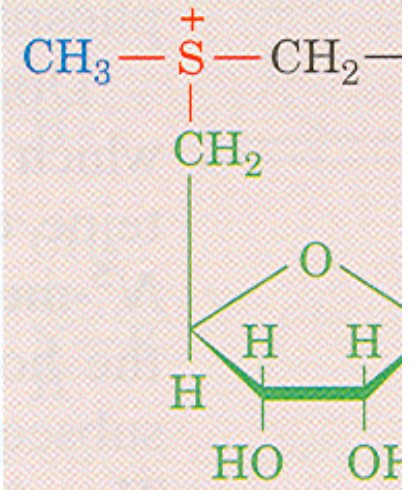
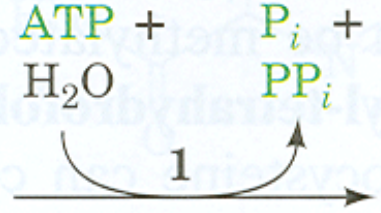


1965 Nobel Pri

François Jacob and Jacques Mo



Methionine

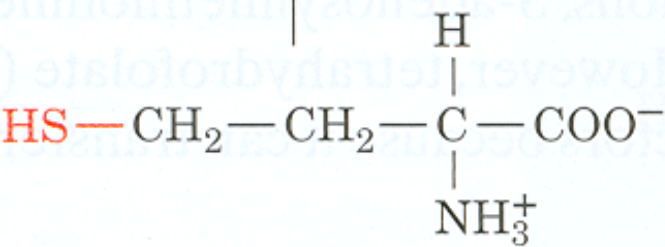


S-Adenosylmethionine

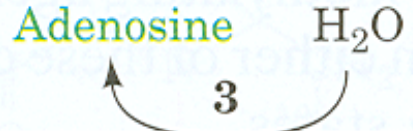


4

THF
N⁵-methyl-THF

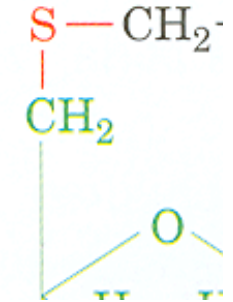


Homocysteine



biosynthetic methylation

2

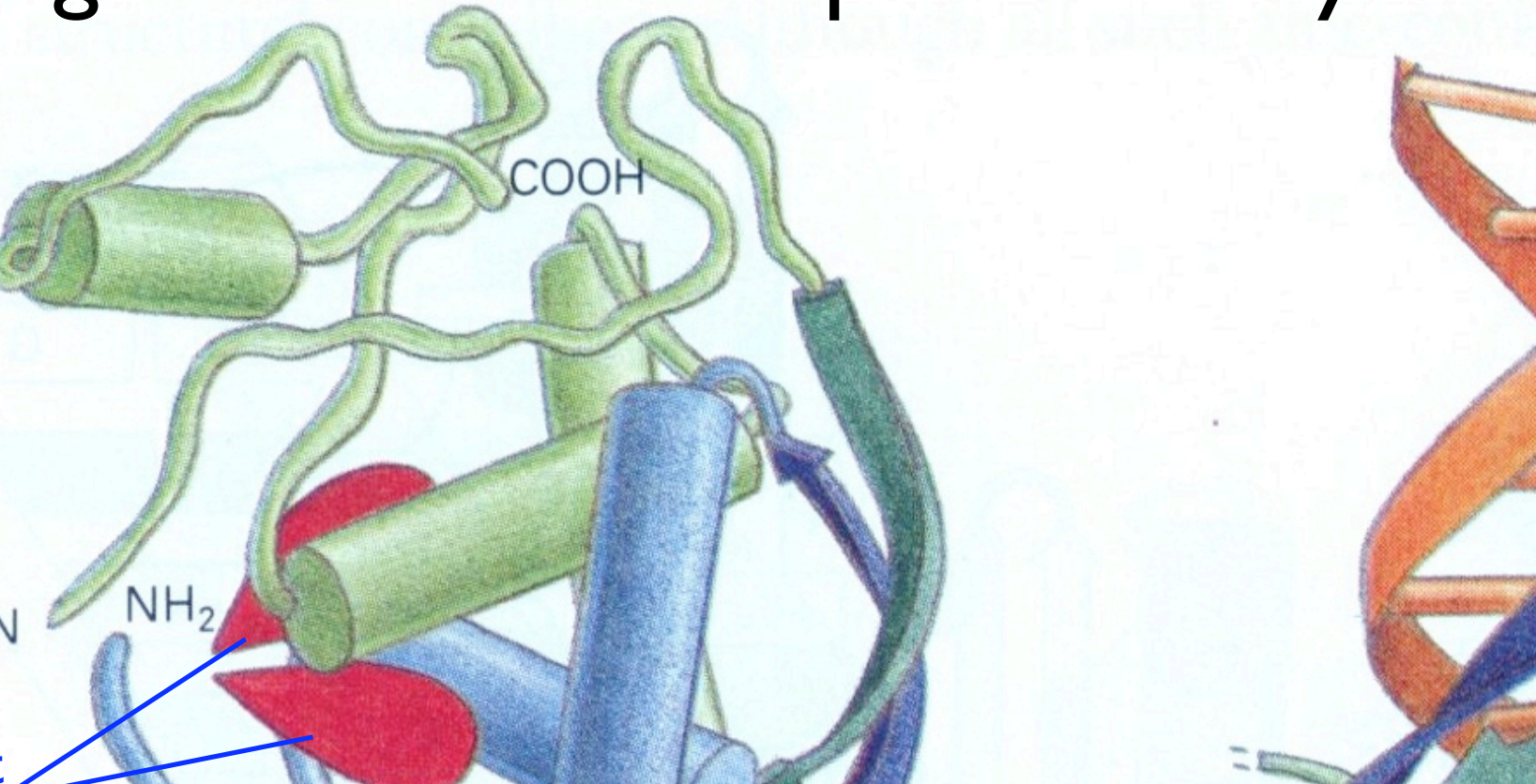


Bacterial Met Repressor

a beta-sheet DNA binding domain

Negative feedback loop:

high Met level \Rightarrow repress Met synthesis



Summary

Proteins can bind DNA to regulate expression (i.e., production of other & themselves)

This is widespread

Complex combinatorial control is p

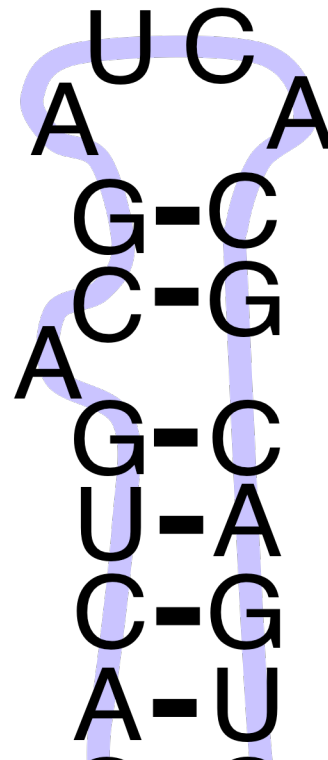
RNA Secondary Structure:

RNA makes helices too

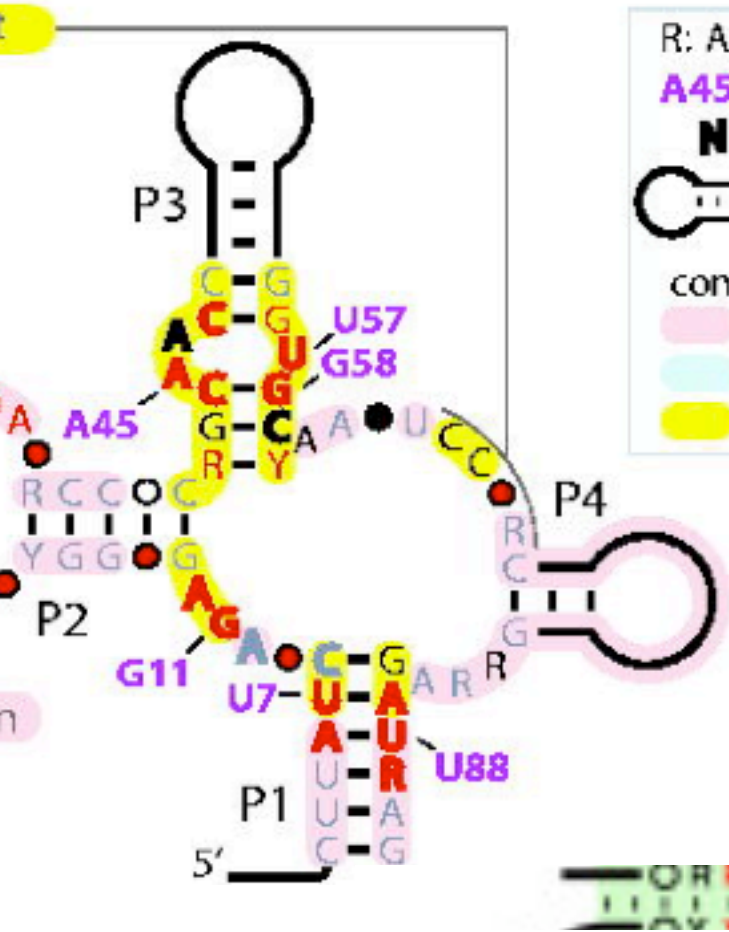
Base pairs

A-U

C-G



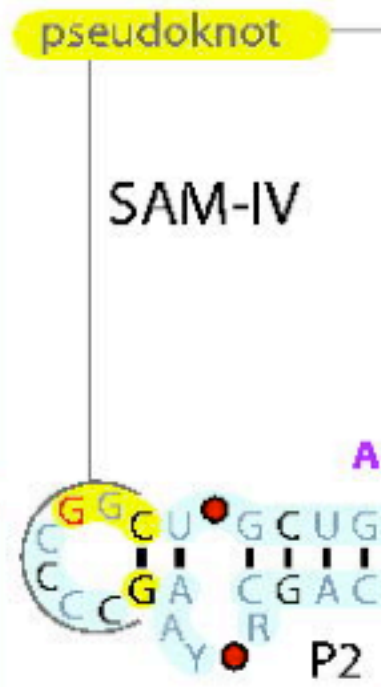
Look, Ma, no pro'

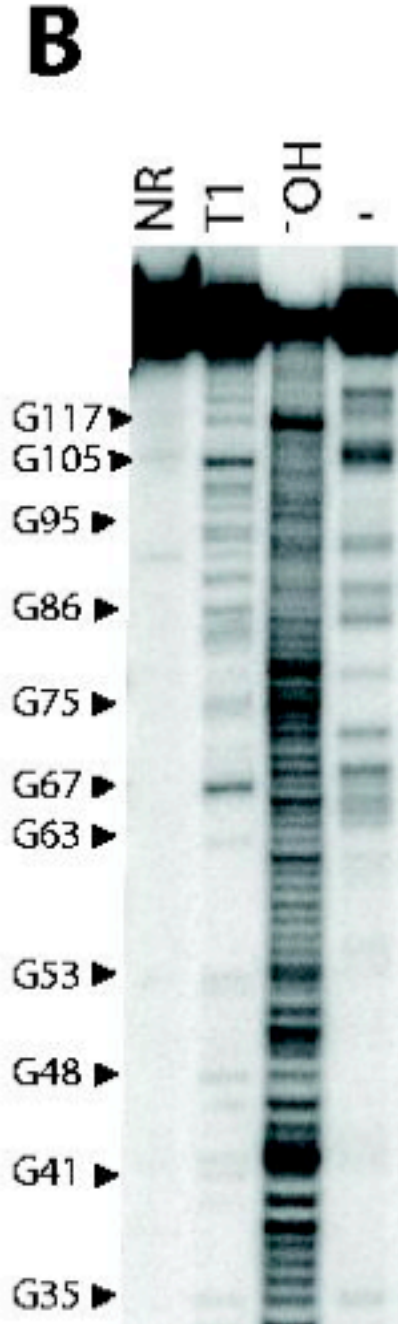
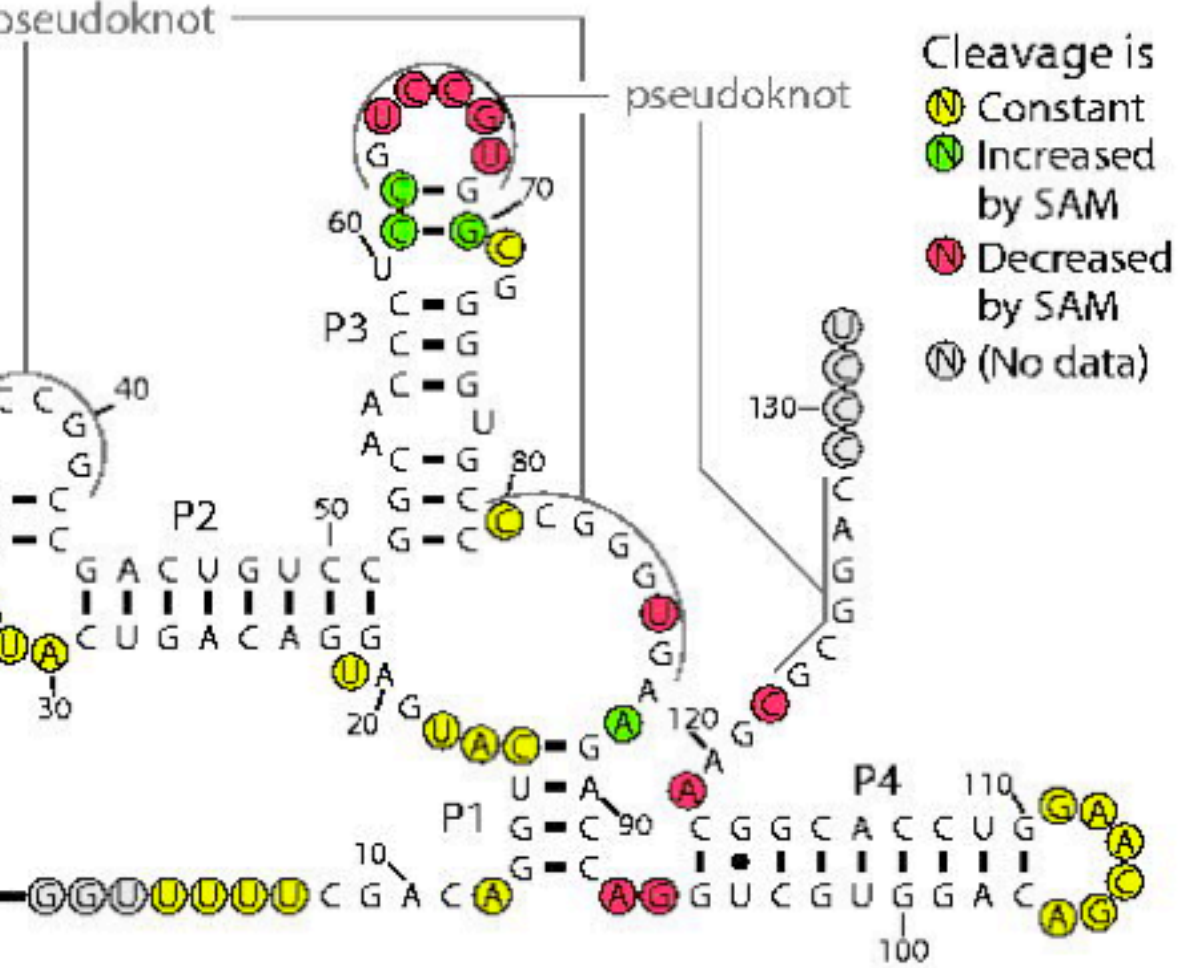


R: A or G. Y: C or U.
A45 position in SAM-I structure
N position in 5 Å shell of ligand
 variable-length hairpin

conserved features ...
 ...unique to SAM-I
 ...unique to SAM-IV
 ...common to both

	nucleotide identity	nucleotide present
N	97%	97%
N	90%	90%
N	75%	75%
		50%





SAM-IV Gene Context

Seq accession		5' at	3' at	genes
003450.3	+	1372526	1372970	RNA → hypo →
006958.1	+	1373993	1374437	RNA → hypo →
004369.1	+	1474341	1474771	RNA → hypo →
008711.1	-	3791869	3791425	RNA → Nitrilotriacetate monooxygenase (cd01095) → Nit DdpA (COG0747) → DppB (COG0601) → DppC (COG
008711.1	-	1118288	1117855	RNA → hypo →
008726.1	-	1175546	1175106	RNA → DSPc (pfam00782)ADP_ribosyl_GH (pfam0374
008596.1	-	1362007	1361570	RNA → DSPc (pfam00782)ADP_ribosyl_GH (pfam0374
003155.3	+	3688850	3689305	RNA → hypo → hypo →
003155.3	-	3688791	3688366	RNA → hypo →
003155.3	+	8585278	8585711	RNA → HisM (COG0765) → GlnQ (COG1126) → SBP
008278.1	+	4609186	4609626	RNA → CsdB (COG0520) →
Y01218155.1	+	112	544	RNA → unknown →
003888.3	-	2308784	2308334	RNA → CsdB (COG0520) →
003155.3	+	7291219	7291670	RNA → CsdB (COG0520) →
AAMN01000002.1	+	104068	104500	RNA → metA (COG2021) →
008699.1	+	3644112	3644553	RNA → CsdB (COG0520) →
008268.1	-	6724436	6724002	RNA → metC (COG2873) → metA (COG2021) →
008711.1	+	1618323	1618754	RNA → metA (COG2021) →
008268.1	+	4627879	4628337	RNA → CsdB (COG0520) →
AAMN01000002.1	+	4561	5010	RNA → CsdB (COG0520) → DegV (COG1307) → DUF
006361.1	-	1024561	1024137	RNA → metA (COG2021) →
AAEF02000064.1	+	13374	13810	RNA → DszC (cd01163) →
008541.1	+	1479272	1479711	RNA → metA (COG2021) → Sugar_tr (pfam00083) → A
008541.1	+	2980393	2980837	RNA → CsdB (COG0520) → SmtA (COG0500) → TroR CrcB (COG0239) → hypo →
008711.1	+	2904748	2905189	RNA → CsdB (COG0520) →
AAMN01000002.1	+	91101	91530	RNA → metC (COG2873) →

NC_008705.1	-	1311762	1311315	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NZ_AAQC01000009.1	+	7131	7578	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NC_008611.1	-	1553444	1552996	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NC_002755.2	+	3723565	3724013	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NC_000962.2	+	3725957	3726405	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NC_002945.3	+	3683207	3683655	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NZ_AAIX01000036.1	+	18996	19444	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NZ_AAKR01000147.1	+	6245	6693	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NZ_AASN01000046.1	+	389912	390360	RNA→					
NC_002944.2	+	3838558	3839009	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NC_008595.1	+	4445969	4446420	RNA→	metC (COG2873)	→	metA (COG2021)	→	U
NC_008541.1	+	1477689	1478157	RNA→	metC (COG2873)	→			
NZ_AAI01000122.1	+	12974	13413	RNA→	CsdB (COG0520)	→			
NZ_AAAP01003574.1	+	1800	2239	RNA→	COG0520: Selenocysteine lyase	→	CsdB (C		
NZ_AAEF02000003.1	+	44250	44673	RNA→	DdpA (COG0747)	→	DppB (COG0601)	→	I
					Nitrilotriacetate monooxygenase (cd01095)	→			
NZ_AAEF02000019.1	-	36781	36357	RNA→	NlpA (COG1464)	→	AbcC (COG1135)	→	A
NC_006361.1	-	5665654	5665216	RNA→	RHOD_1 (cd01522)	→	metZ (pfam01053)	→	
NZ_AATJ01000006.1	-	318699	318245	RNA→	CsdB (COG0520)	→			
NZ_AAWA01000001.1	-	73920	73461	RNA→	CsdB (COG0520)	→			
NC_006361.1	+	396879	397316	RNA→	CsdB (COG0520)	→			
NZ_AAEF02000050.1	-	6285	5844	RNA→	CsdB (COG0520)	→			
NC_008278.1	+	4831332	4831773	RNA→	DdpA (COG0747)	→	DppB (COG0601)	→	
NC_008541.1	+	4342124	4342550	RNA→	COG5515 (COG5515)	→	RHOD_1 (cd01522)		
NC_008711.1	+	4015346	4015771	RNA→	RHOD_1 (cd01522)	→	metZ (pfam01053)	→	
NZ_AAGP01000018.1	-	74585	74159	RNA→	hypo	→			
NZ_AAQC01000005.1	+	250651	251094	RNA→	←ABC_Class3 (cd03229)				

Supplementary Figure S3C: conserved domains present in genes downstream

domains found in downstream genes (Supplementary Figure S3B) with the first sentence in their description from the Conserved Domain Database. Conserved domains downstream of more than one SAM-IV riboswitch

acetate monooxygenase oxidizes nitrilotriacetate utilizing reduced flavin mononucleotide (FMN) and oxygen.

2-mercaptopyruvate sulfurtransferase (DBT) desulfurization enzyme C (DszC).

Member of the Rhodanese Homology Domain superfamily, subgroup 1.

This class is comprised of all BPD (Binding Protein Dependent) systems that are largely conserved in archaea and eubacteria and are primarily involved in scavenging solutes from the environment.

Integral membrane protein possibly involved in chromosome condensation [Cell division and chromosome partitioning]

Secondary independent methyltransferases [Secondary metabolites biosynthesis, transport, and regulation] / General function prediction only]

Cysteine lyase [Amino acid transport and metabolism]

ABC-type dipeptide/oligopeptide/nickel transport systems, permease components [Amino acid transport and metabolism / Inorganic ion transport and metabolism]

Cystathionine beta-lyases/cystathionine gamma-synthases [Amino acid transport and metabolism]

ATPase-like ATPases [General function prediction only]

ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism]

ABC-type amino acid transport system, permease component [Amino acid transport and metabolism]

are assigned a color, while others are copied from supplementary data on a domain-by-domain basis.

transport and metabolism]

COG1135 ABC-type metal ion transport system and metabolism]

COG1173 ABC-type dipeptide/oligopeptide transport system [Amino acid transport and metabolism / Inorganic ion transport and metabolism]

COG1307 Uncharacterized protein conserved in archaea

COG1321 Mn-dependent transcriptional regulator

COG1464 ABC-type metal ion transport system and metabolism]

COG1715 Restriction endonuclease [Defense mechanisms]

COG2011 ABC-type metal ion transport system and metabolism]

COG2021 Homoserine acetyltransferase [Amino acid transport and metabolism]

COG2226 Methylase involved in ubiquinone biosynthesis

COG2873 O-acetylhomoserine sulfhydrylase

COG3393 Predicted acetyltransferase [General function prediction only]

COG4268 McrBC 5-methylcytosine restriction endonuclease

COG5515 Uncharacterized conserved small protein

pfam00083 Sugar (and other) transporter.

pfam00497 Bacterial extracellular solute-binding protein

pfam00782 Dual specificity phosphatase, cytoplasmic

pfam01053 Cys/Met metabolism PLP-dependent