

CSE 427

Computational Biology

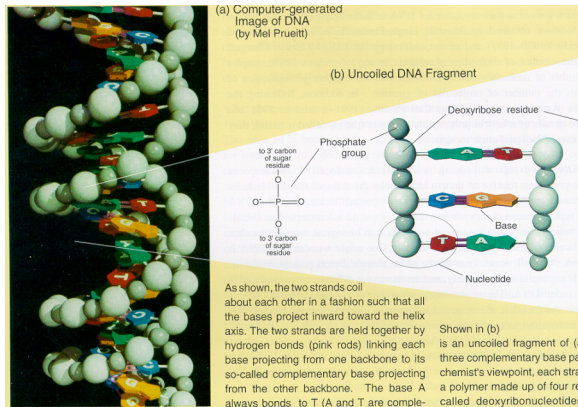
Gene Regulation
L. Ruzzo
Winter 2008

DNA Binding Proteins

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

2

The Double Helix

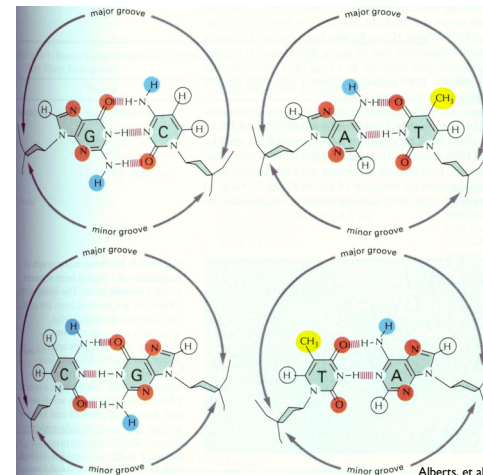


Los Alamos Science

3

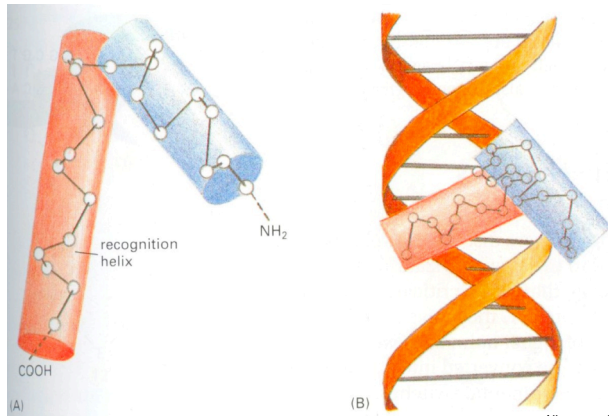
In the groove

Different patterns of potential H bonds at edges of different base pairs, accessible esp. in major groove

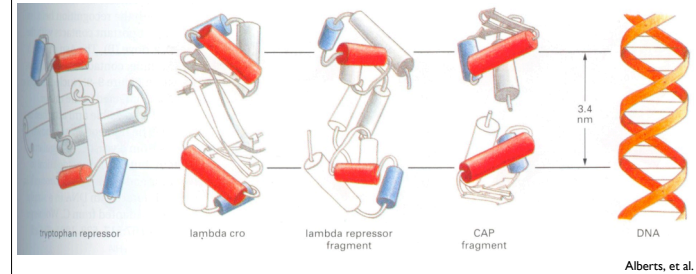


4

Helix-Turn-Helix DNA Binding Motif



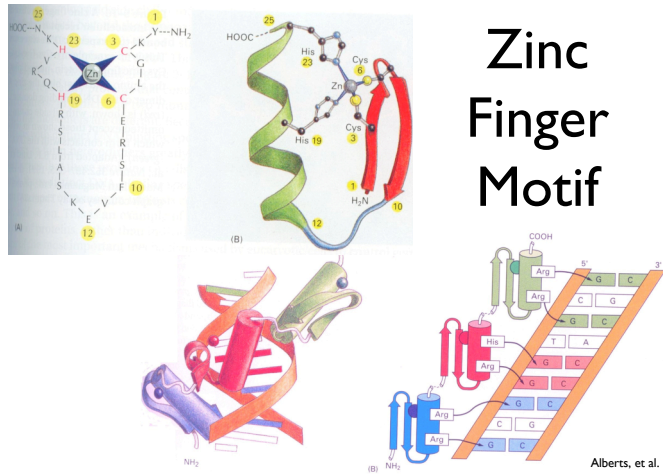
H-T-H Dimers



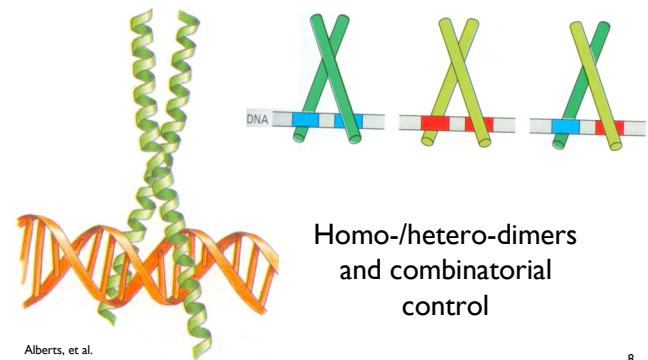
Bind 2 DNA patches, ~ 1 turn apart
Increases both specificity and affinity

6

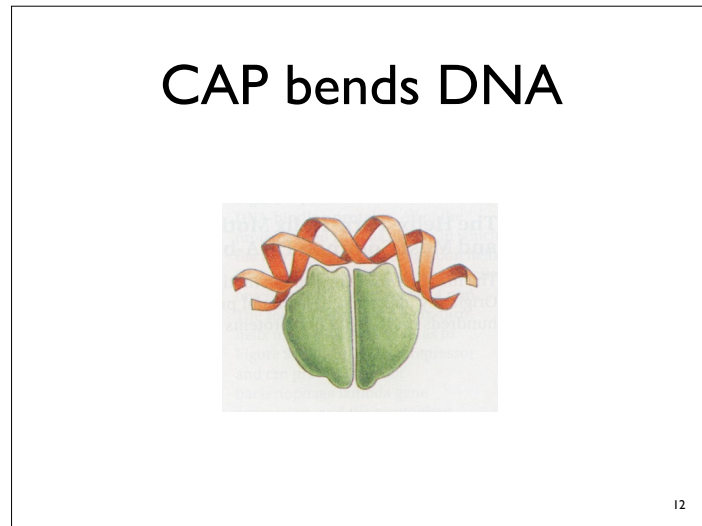
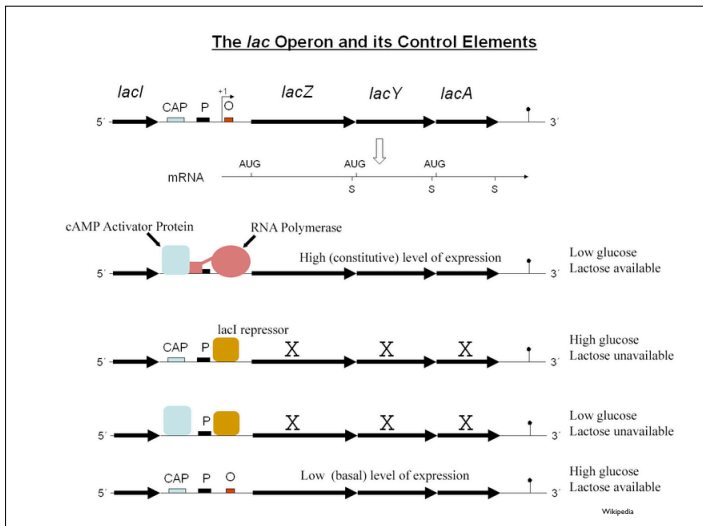
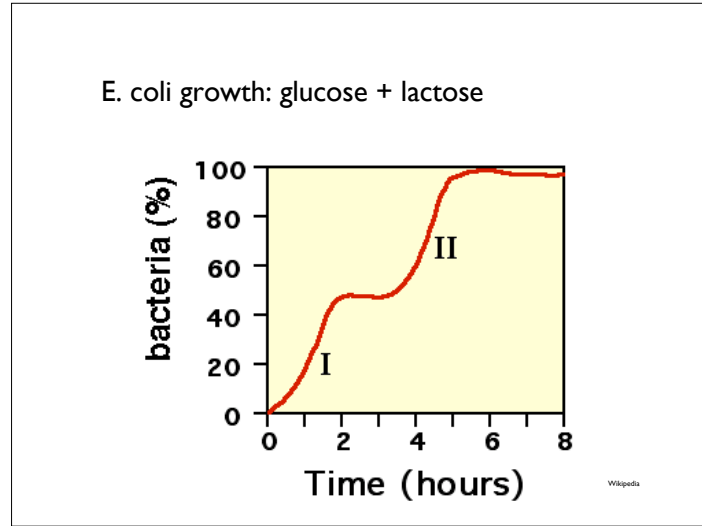
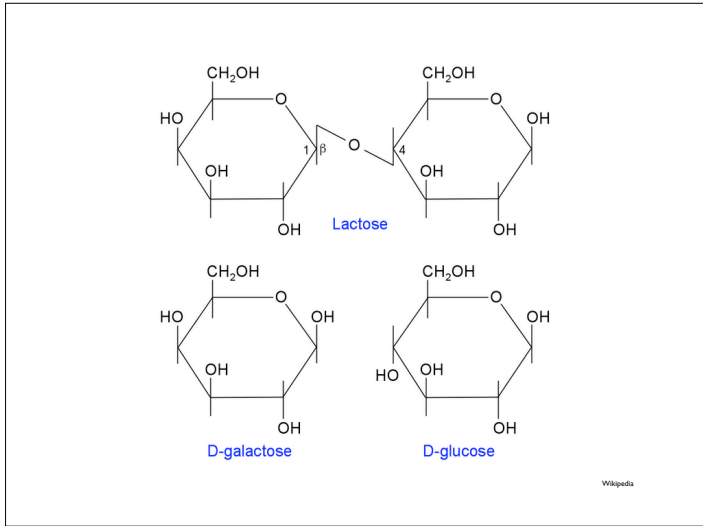
Zinc Finger Motif



Leucine Zipper Motif



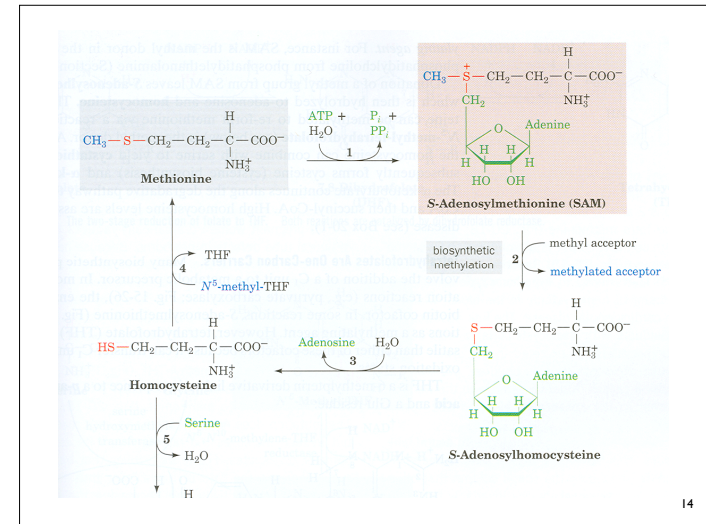
8



1965 Nobel Prize

François Jacob and Jacques Monod

13



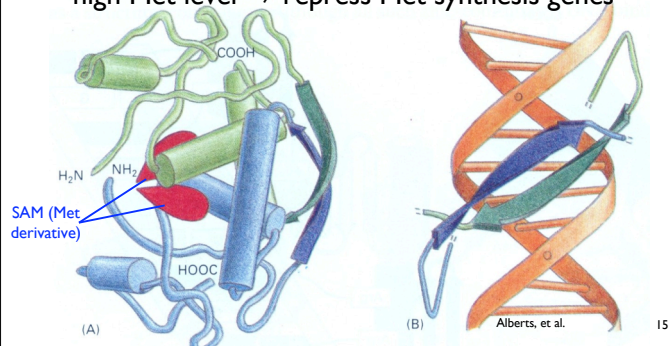
14

Bacterial Met Repressor

a beta-sheet DNA binding domain

Negative feedback loop:

high Met level ⇒ repress Met synthesis genes



Summary

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

This is widespread

Complex combinatorial control is possible

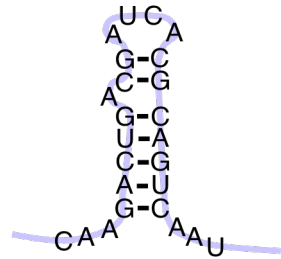
But it's not the only way to do this...

16

RNA Secondary Structure: RNA makes helices too

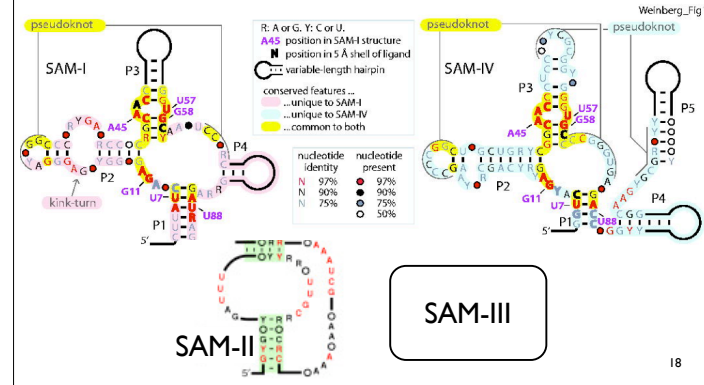
Base pairs

A-U
C-G

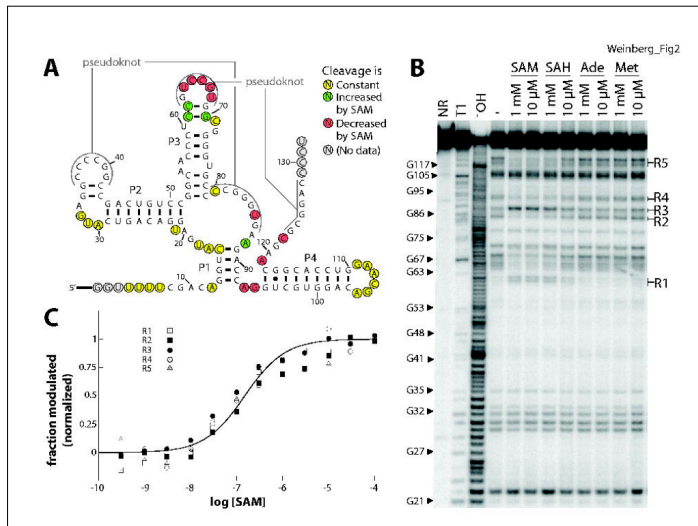


17

Look, Ma, no proteins



18



SAM-IV Gene Context

abbrev	RefSeq accession	5' at	3' at	genes
Cgl-1-1	NC_003450.3	+	1372526	1372970 RNA - hypo-
Cgl-1-2	NC_006958.1	+	1378993	1374437 RNA - hypo-
Cgl-1-1	NC_004689.1	+	1474341	1474771 RNA - hypo-
Aas-1-1	NC_008711.1	-	3791869	3791425 RNA - Nitrilotriacetate monooxygenase (cd1089) - DdpA (COG0747) - DdpB (COG0601) - DdpC (COG1173) - COG1123 (COG1123) - RNA - hypo-
Aas-1-2	NC_008711.1	-	1118288	1117855 RNA - hypo-
Mvm-1-1	NC_008726.1	-	1175546	1175106 RNA - DSPc (pfam00782)ADP-ribosyl.GH (pfam03747) - hypo-
Mam-1-1	NC_008596.1	-	1362007	1361570 RNA - DSPc (pfam00782)ADP-ribosyl.GH (pfam03747) - hypo-
Sav-1-1	NC_003155.3	+	3688850	3688905 RNA - hypo- - hypo-
Sav-1-2	NC_003155.3	+	3688791	3688366 RNA - hypo-
Sav-1-3	NC_003155.3	+	8585278	8585711 RNA - HisM (COG0765) - GlnQ (COG1126) - SHP_bac.3 (pfam04097) - COG3893 (COG3893) -
Fal-1-1	NC_009278.1	+	4609186	4609626 RNA - CsdB (COG0520) -
env-1	AACY01218155.1	+	112	544 RNA - unknown -
Seo-1-1	NC_008888.3	+	2308784	2308334 RNA - CsdB (COG0520) -
Sav-1-4	NC_003155.3	+	7291219	7291670 RNA - CsdB (COG0520) -
Jap-1-1	NZ_AAMN01000002.1	+	104098	104500 RNA - metA (COG2021) -
Npp-1-1	NC_008699.1	+	3644112	3644553 RNA - CsdB (COG0520) -
Rop-1-1	NC_008698.1	-	6724436	6724202 RNA - metC (COG2028) - metA (COG2021) -
Aas-1-3	NC_008711.1	+	1618323	1618754 RNA - metA (COG2021) -
Rap-1-2	NC_008268.1	+	4627879	4628337 RNA - CsdB (COG0520) -
Jap-1-2	NZ_AAMN01000002.1	+	4561	5010 RNA - CsdB (COG0520) - DegV (COG1307) - DUF205 (pfam2660) -
Ntr-1-1	NC_008961.1	-	1034561	1034197 RNA - metA (COG2021) -
Kra-1-1	NZ_AAEF02000084.1	+	13374	13810 RNA - DnaC (cd01183) -
App-1-1	NC_008541.1	+	1479272	1479711 RNA - metA (COG2021) - SiguarA (pfam0088) - AzfC (COG1092) -
App-1-2	NC_008541.1	+	2980393	2980657 RNA - metA (COG2021) - TrnH (COG1321)IpsxA (pfam04028) - CrcB (COG2028) - hypo-
Aas-1-4	NC_008711.1	+	2904748	2905189 RNA - CsdB (COG0520) -
Jap-1-3	NZ_AAMN01000002.1	+	91101	91530 RNA - metC (COG2028) -
Mle-1-1	NC_002877.1	-	819835	819388 RNA - metA (COG2021) -
Mam-1-2	NC_008696.1	+	1744807	1747949 RNA - metC (COG2028) - metA (COG2021) -
MB-1-1	NZ_AAP01000001.1	+	833500	833951 RNA - metC (COG2028) - metA (COG2021) -
Mva-1-2	NC_008726.1	-	1643759	1643310 RNA - metC (COG2028) - metA (COG2021) -
Map-3-1	NC_008146.1	-	1307453	1307006 RNA - metC (COG2028) - metA (COG2021) -

20

Msp-2-1	NC.008705.1	-	1311762	1311315	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Msp-1-1	NZ_AAQC01000009.1	+	7131	7078	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mtu-1-1	NC.008611.1	-	1533444	152996	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mtu-2-1	NC.002755.2	+	3723565	3724013	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mtu-4-1	NC.000962.2	+	3729587	3726405	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mbo-1-1	NC.002845.3	+	3683207	3683655	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mtu-3-1	NZ_AADK01000036.1	+	18996	19444	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mtu-1-1	NZ_AAKR01000147.1	+	6245	6663	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mtu-2-1	NZ_AASS01000046.1	+	389912	390360	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mav-2-1	NC.002944.2	+	3838558	3839009	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Mav-1-1	NC.008595.1	+	4445969	4446420	RNA → hscC (COG2873) → hstA (COG2021) → hscC (COG2873) →
Asp-1-3	NC.008541.1	+	1477689	1478157	RNA → hscC (COG2873) →
Psp-1-1	NZ_AAH01000122.1	+	12974	13413	RNA → CsdB (COG0520) →
Mma-1-1	NZ_AAAP01003574.1	+	1800	2239	RNA → COG0520: Selenocysteine lyase → CsdB (COG0520) →
*Kra-1-2	NZ_AAEF02000003.1	+	44250	44673	RNA → DtpA (COG0747) → DppB (COG0601) → DppC (COG1178) → COG1124 (COG1129) → NtpA (COG1464) → NtpB (COG1465) →
Kra-1-3	NZ_AAEF020000019.1	-	36781	36357	RNA → NtpA (COG1464) → AbcC (COG1135) → AbcD (COG2011) →
Nta-1-2	NC.006061.1	-	566564	565216	RNA → RHO2-1 (cd01322) → AbcC (COG1135) →
Str-1-1	NZ_AATJ01000006.1	-	318699	318245	RNA → CsdB (COG0520) →
Str-1-1	NZ_AAWA01000001.1	-	73920	73461	RNA → CsdB (COG0520) →
Nta-1-3	NC.00630.1	+	396879	397316	RNA → CsdB (COG0520) →
Kra-1-4	NZ_AAEF020000050.1	+	6285	5844	RNA → CsdB (COG0520) →
Pal-1-2	NC.008278.1	+	4831332	4831773	RNA → DtpA (COG0747) → DppB (COG0601) →
*Asp-1-4	NC.00841.1	+	434124	434250	RNA → COG5519 (COG5513) → RHO2-1 (cd01322) → MerC (COG0926) →
Asu-1-5	NC.008711.1	+	4013346	4013771	RNA → RHO2-1 (cd01322) → hscC (COG2873) → putative integral membrane protein →
Blu-1-1	NZ_AAGP01000018.1	+	74585	74159	RNA → hspB →
Msp-1-2	NZ_AAQC01000005.1	+	250651	251094	RNA → --ABC Class0 (cd03229) →
MB-1-2	NZ_AFA01000002.1	+	124363	124836	RNA → ABC3 (pfam0793) →
Msp-3-2	NC.008146.1	-	920185	919752	RNA → hspB →
Msp-2-2	NC.008705.1	-	928924	928481	RNA → hspB → hspB → COG0714 (COG0714)Mr (COG1175) → MerC (COG4268) →

21

Supplementary Figure S3C: conserved domains present in genes downstream of SAM-IV riboswitches

Conserved domains found in downstream genes (Supplementary Figure S3B) are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains downstream of more than one SAM-IV riboswitch are assigned a color, while others are shown in gray. (This explanatory text is largely copied from supplementary data on a different RNA motif [4].)

000088	nitrilotriacetate monooxygenase oxidase/nitrilotriacetate-utilizing reduced flavin monooxidase (FAMHD) and oxygen	
000186	Dihydrodipicolinate (DHP) identification enzyme C (DhcC)	
000322	Member of the Helicase Homology Domain superfamily, subgroup 1.	
000229	This class is comprised of all HPD (Helicase Protein Dependent) systems that are largely represented in archaea and eubacteria and are primarily involved in scavenging acetate from the environment.	
000269	Integral membrane protein possibly involved in chromosome condensation / Cell division and chromosome partitioning.	
COG000	SAM-dependent methyltransferase [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]	
COG001	Selenocysteine lyase [Amino acid transport and metabolism]	
COG002	ABC-type dipeptide/oligopeptide/peptide transport systems, permease components [Amino acid transport and metabolism / Inorganic ion transport and metabolism]	
COG006	Cystathionine beta-lyase/cystathionine gamma-synthase [Amino acid transport and metabolism]	
COG007	ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism]	
COG008	MinD-like ATPase [General function prediction only]	
COG009	ABC-type dipeptide transport system, permease component [Amino acid transport and metabolism]	
COG002	Zinc-dependent alcohol dehydrogenase, class III [Energy production and conversion]	
COG010	ATPase components of various ABC-type transport systems, contain duplicated ATPase [General function prediction only]	
COG106	ABC-type polyanion transport system, ATPase component [Amino acid transport and metabolism]	
COG118	ABC-type metal ion transport system, ATPase component [Inorganic ion transport and metabolism]	
COG119	ABC-type dipeptide/oligopeptide/peptide transport systems, permease component [Amino acid transport and metabolism / Inorganic ion transport and metabolism]	
COG180	Uncharacterized protein conserved in bacteria [Function unknown]	
COG182	Mn-dependent transcriptional regulator [Transcription]	
COG184	ABC-type metal ion transport system, periplasmic component/surface antigen [Inorganic ion transport and metabolism]	
COG121	Restriction endonuclease [Defense mechanism]	
COG201	ABC-type metal ion transport system, permease component [Inorganic ion transport and metabolism]	
COG202	Isomerase acetyltransferase [Amino acid transport and metabolism]	
COG203	Methylase involved in ubiquitin/ubiquitin-like biosynthesis [Conjugase metabolite]	
COG204	O-acetylserine sulfurylase [Amino acid transport and metabolism]	
COG205	Predicted acetyltransferase [General function prediction only]	
COG206	O-acetylserine sulfurylase restriction system component [Defense mechanism]	
COG208	Uncharacterized conserved small protein [Function unknown]	
pfam00083	Sugar (and other) transporter.	
pfam00087	Essential extracellular matrix-binding protein, family 3.	
pfam00292	Dual specificity phosphatase, catalytic domain.	
pfam00700	Cys/Met metabolism PLP-dependent enzyme.	
pfam02702	Domain of unknown function DUF.	
pfam02732	ENCCA domain.	
pfam03007	ADN-ribosyl/tyrosinohydrolase.	
pfam03023	FovA domain.	

22