Single-letter and three-letter symbols for amino acids

Amino acid	One letter	Three letter
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val
Unspecified or unknown	X	Xaa

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

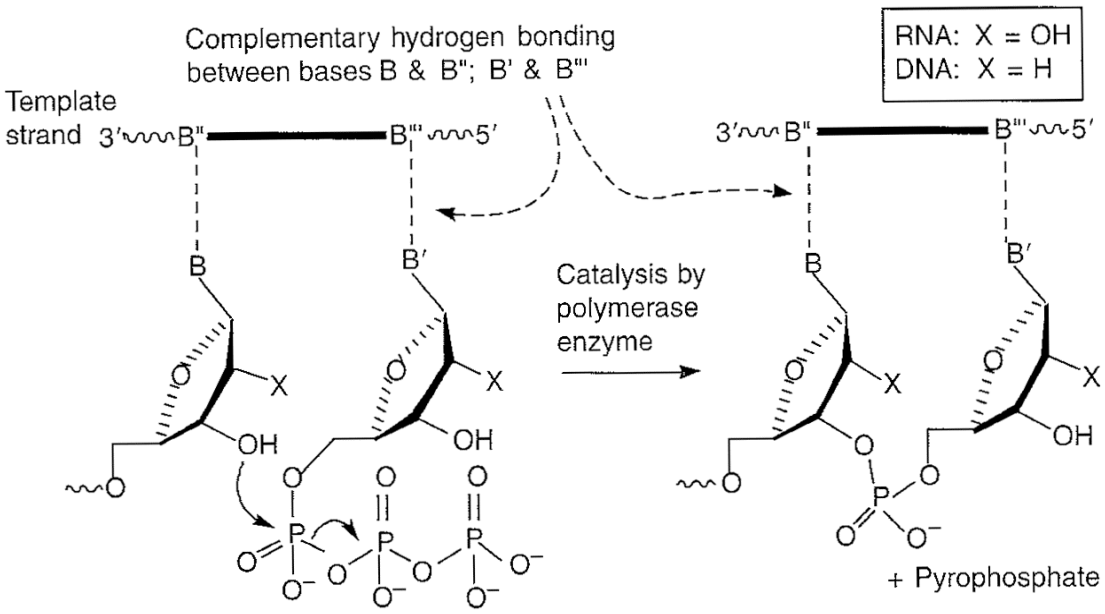
5′ base	Middle base U	Middle base C	Middle base A	Middle base G	3′ base
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA Leu	UCA Ser	UAA Stop*	UGA Stop*	**A**
	UUG Leu	UCG Ser	UAG Stop*	UGG Trp	**G**
C	CUU Leu	CCU Pro	CAU His	CGU Arg	**U**
	CUC Leu	CCC Pro	CAC His	CGC 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**

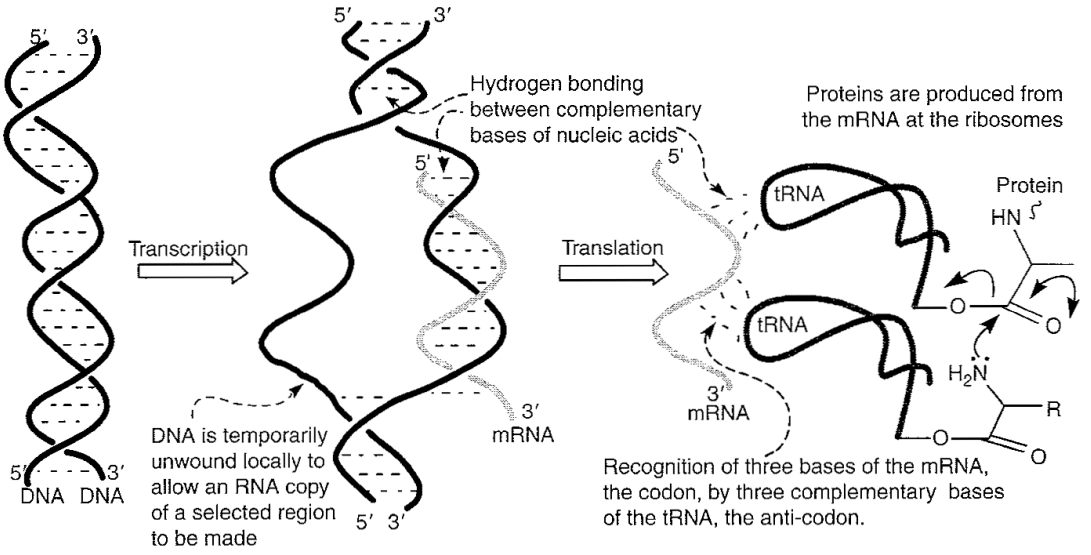
*Stop codons have no amino acids assigned to them.

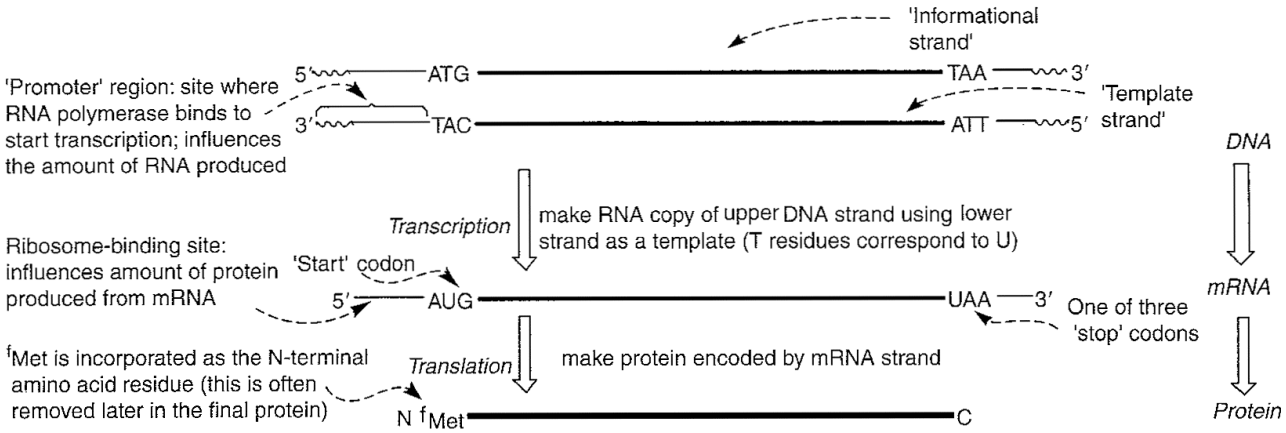
†The AUG codon is the usual initiation codon as well as that for methionine residues elsewhere. The code is almost universal but differences have been found in mitochondrial DNA from some organisms

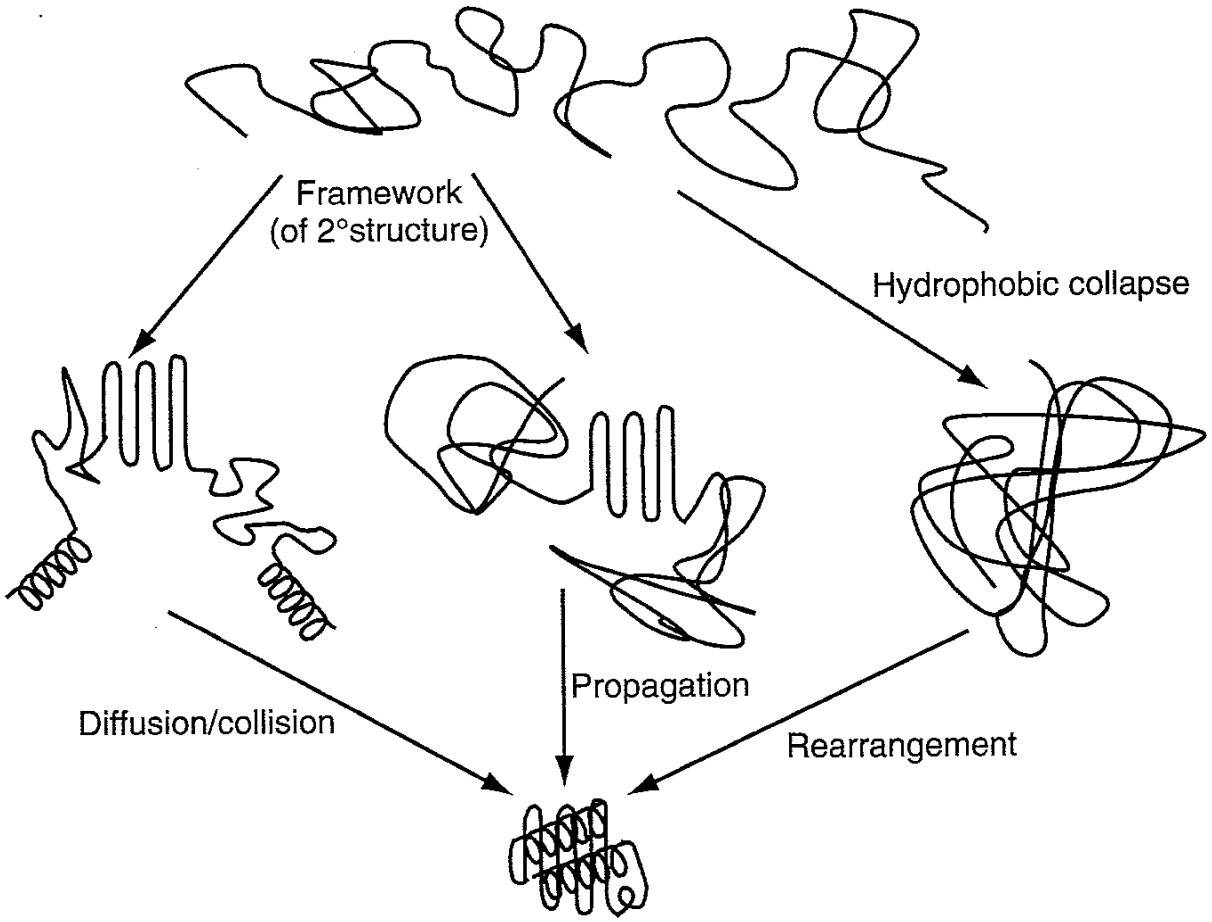
Genome sizes

Organism	*Number of Base pairs*	*Number of Genes*	*Comment*
ϕX–174	5 386	10	virus infecting *E. coli*
Human mitochondrion	16 569	37	subcellular organelle
Epstein–Barr virus (EBV)	172 282	80	cause of mononucleosis
Mycoplasma pneumoniae	816 394	680	cause of cyclic pneumonia epidemics
Rickettsia prowazekii	1 111 523	878	bacterium, cause of epidemic typhus
Treponema pallidum	1 138 011	1 039	bacterium, cause of syphilis
Borrelia burgdorferi	1 471 725	1 738	bacterium, cause of Lyme disease
Aquifex aeolicus	1 551 335	1 749	bacterium from hot spring
Thermoplasma acidophilum	1 564 905	1 509	archaeal prokaryote, lacks cell wall
Campylobacter jejuni	1 641 481	1 708	frequent cause of food poisoning
Helicobacter pylori	1 667 867	1 589	chief cause of stomach ulcers
Methanococcus jannaschii	1 664 970	1 783	archaeal prokaryote, thermophile
Hemophilus influenzae	1 830 138	1 738	bacterium, cause of middle ear infections
Thermotoga maritima	1 860 725	1 879	marine bacterium
Archaeoglobus fulgidus	2 178 400	2 437	another archaeon
Deinococcus radiodurans	3 284 156	3 187	radiation-resistant bacterium
Synechocystis	3 573 470	4 003	cyanobacterium, 'blue-green alga'
Vibrio cholerae	4 033 460	3 890	cause of cholera
Mycobacterium tuberculosis	4 411 529	4 275	cause of tuberculosis
Bacillus subtilis	4 214 814	4 779	popular in molecular biology
Escherichia coli	4 639 221	4 406	molecular biologists' all-time favourite
Pseudomonas aeruginosa	6 264 403	5 570	largest prokaryote sequenced as yet
Saccharomyces cerevisiae	12.1×10^6	5 885	yeast, first eukaryotic genome sequenced
Caenorhabditis elegans	95.5×10^6	19 099	the worm
Arabidopsis thaliana	1.17×10^8	25 498	flowering plant (angiosperm)
Drosophila melanogaster	1.8×10^8	13 601	the fruit fly
Fugu rubripes	3.9×10^8	30 000	puffer fish (fugu fish)
Human	3.2×10^9	34 000?	
Wheat	16×10^9	30 000	
Salamander	10^{11}	?	
Psilotum nudum	10^{11}	?	whisk fern – a simple plant


Complementary hydrogen bonding
between bases B & B"; B' & B"'
RNA: X = OH
DNA: X = H
Template
strand 3′ B" B"' 5′
Catalysis by
polymerase
enzyme
+ Pyrophosphate


5′ 3′
DNA DNA
Transcription
Hydrogen bonding
between complementary
bases of nucleic acids
DNA is temporarily
unwound locally to
allow an RNA copy
of a selected region
to be made
3′
mRNA
Translation
Proteins are produced from
the mRNA at the ribosomes
tRNA
tRNA
Protein
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



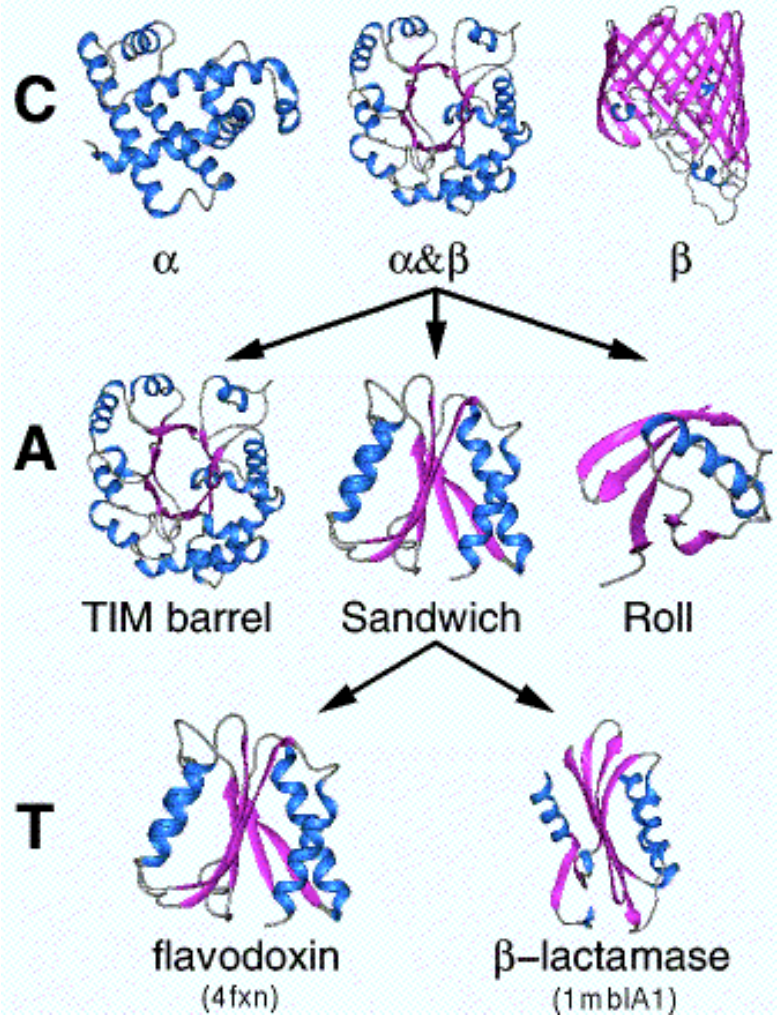
A database for 3D shapes of protein domains

C Class – Secondary Structure

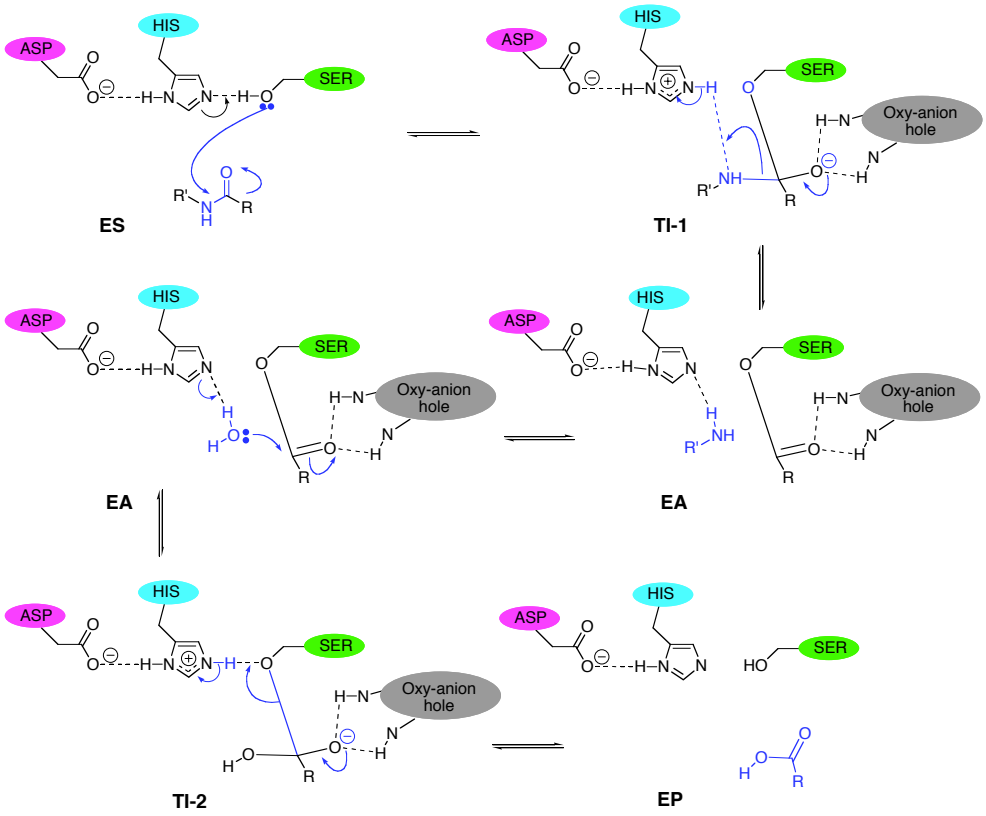
A Architecture – Arrangements of 2ndary Structure but ignores connections

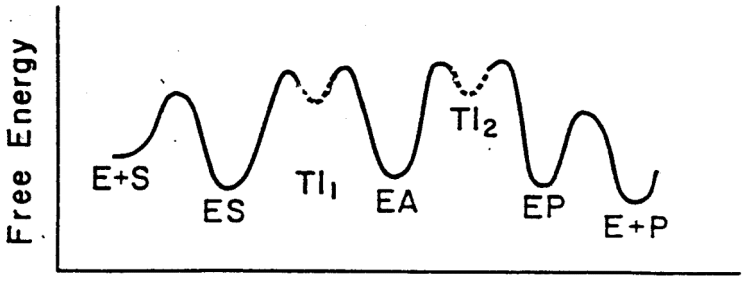
T Topology – Clustering of A

H Homologous Superfamliy – Combines Sequence and Structure information to create groups.



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

Additional Reading:

• Ashburner *et al*; Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **2000**, *25*, 25-29

• Ursing *et al*; EXProt: a database for proteins with experimentally verified function. *Nucleic Acids Research* **2002**, *30*, 50-51.

• Orengo *et al* CATH- A Hierarchic Classification of Protein Domain Structures. *Structure* **1997**, *5*, 1093-1108.

• Pearl *et al* Assigning genomic sequences to CATH *Nucleic Acids Research*. **2000**, *28*, 277-282