

Unit III

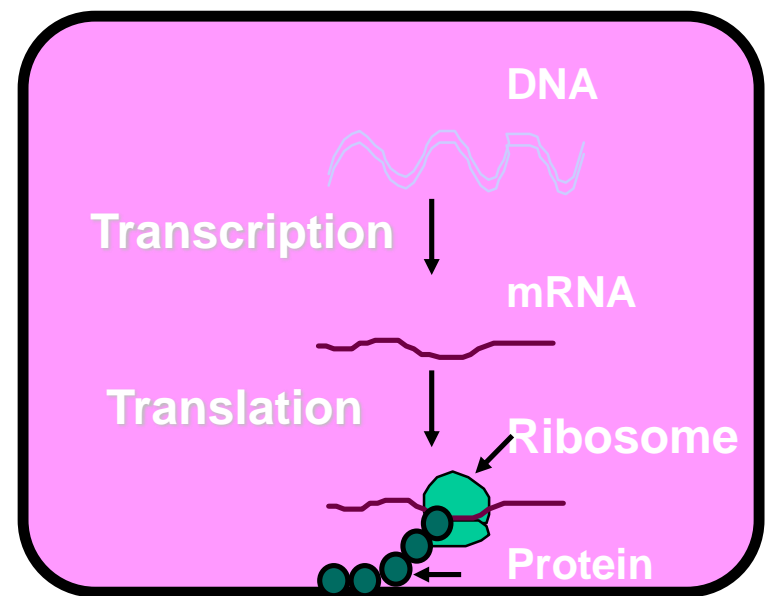
- Protein synthesis
- Secondary structure of the protein
- Structure and function
- Structural databases
- Protein visualizing tools
- Secondary structure prediction algorithms

Protein Synthesis

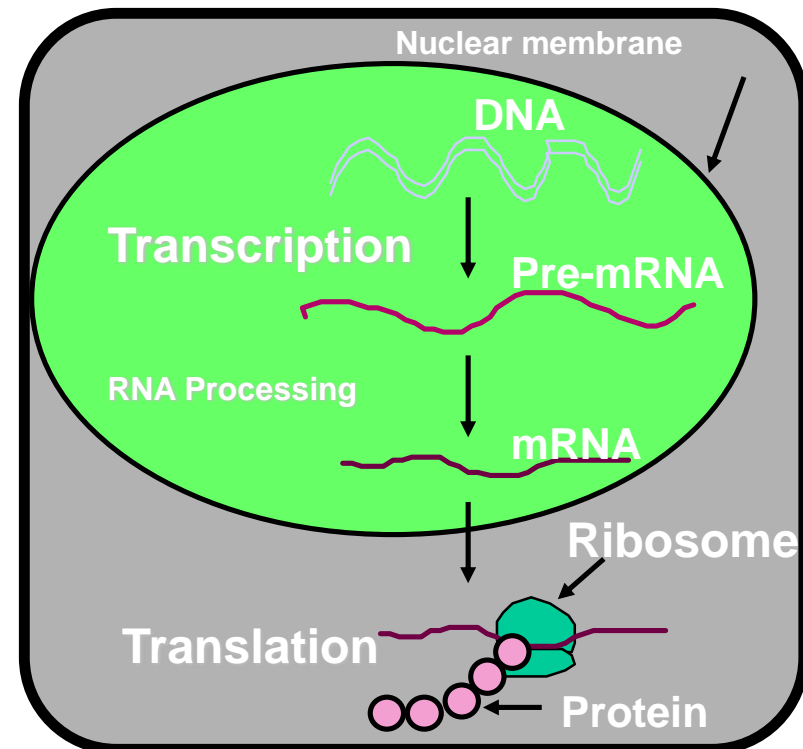
Protein Synthesis

- The production (synthesis) of polypeptide chains (proteins)
- Two phases:
Transcription & Translation
- mRNA must be processed before it leaves the nucleus of eukaryotic cells

Prokaryotic Cell



Eukaryotic Cell



Three Types of RNA

- **Messenger RNA (mRNA)** carries genetic information to the ribosomes
(blueprint for the construction of a protein)
- **Ribosomal RNA (rRNA)**, along with protein, makes up the ribosomes
(construction site where the protein is made)
- **Transfer RNA (tRNA)** transfers amino acids to the ribosomes where proteins are synthesized
(truck delivering the proper amino acid to the site at the right time)

Genes & Proteins

- **Proteins** are made of **amino acids** linked together by **peptide bonds**
- **20** different amino acids **exist**
- Amino acids chains are called **polypeptides**
- Segment of DNA that codes for the amino acid sequence in a protein are called **genes**

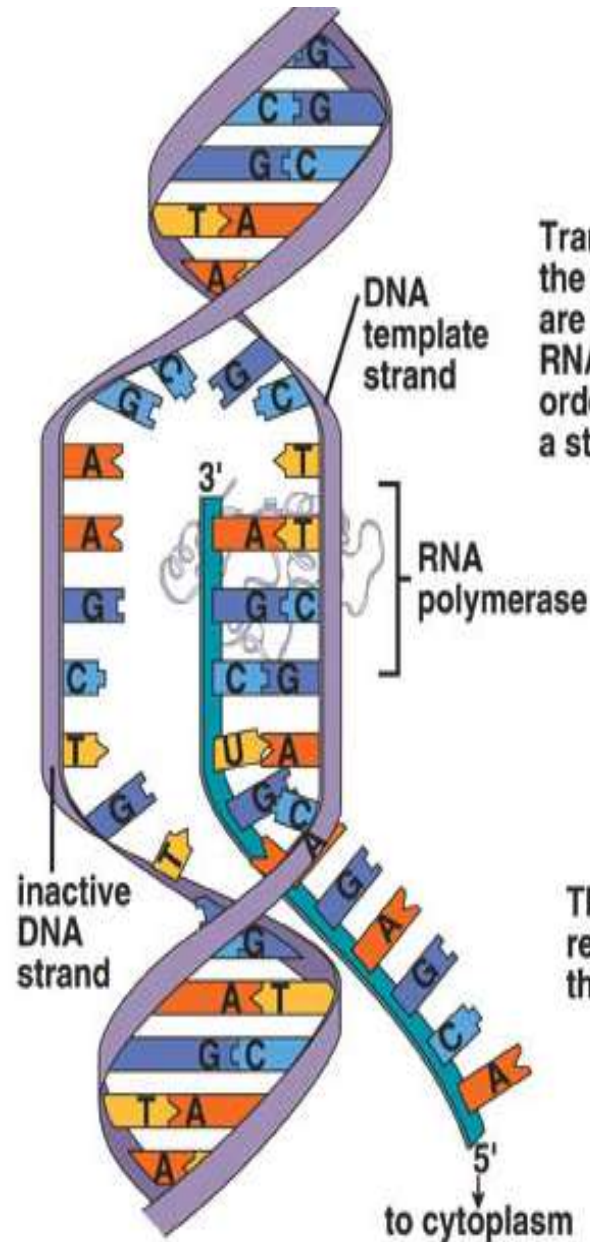
Genetic Code:

- DNA contains a **triplet code**
- Every three bases on DNA stands for **ONE amino acid**
- Each three-letter unit on **mRNA** is called a **codon**
- **Most amino acids have more than one codon!**
- There are **20 amino acids** with a possible 64 different triplets
- The code is nearly **universal** among living organisms

First Base	Second Base				Third Base
	U	C	A	G	
U	UUU phenylalanine	UCU serine	UAU tyrosine	UGU cysteine	U
	UUC phenylalanine	UCC serine	UAC tyrosine	UGC cysteine	C
	UUA leucine	UCA serine	UAA stop	UGA stop	A
	UUG leucine	UCG serine	UAG stop	UGG tryptophan	G
C	CUU leucine	CCU proline	CAU histidine	CGU arginine	U
	CUC leucine	CCC proline	CAC histidine	CGC arginine	C
	CUA leucine	CCA proline	CAA glutamine	CGA arginine	A
	CUG leucine	CCG proline	CAG glutamine	CGG arginine	G
A	AUU isoleucine	ACU threonine	AAU asparagine	AGU serine	U
	AUC isoleucine	ACC threonine	AAC asparagine	AGC serine	C
	AUA isoleucine	ACA threonine	AAA lysine	AGA arginine	A
	AUG (<i>start</i>) methionine	ACG threonine	AAG lysine	AGG arginine	G
G	GUU valine	GCU alanine	GAU aspartate	GGU glycine	U
	GUC valine	GCC alanine	GAC aspartate	GGC glycine	C
	GUA valine	GCA alanine	GAA glutamate	GGA glycine	A
	GUG valine	GCG alanine	GAG glutamate	GGG glycine	G

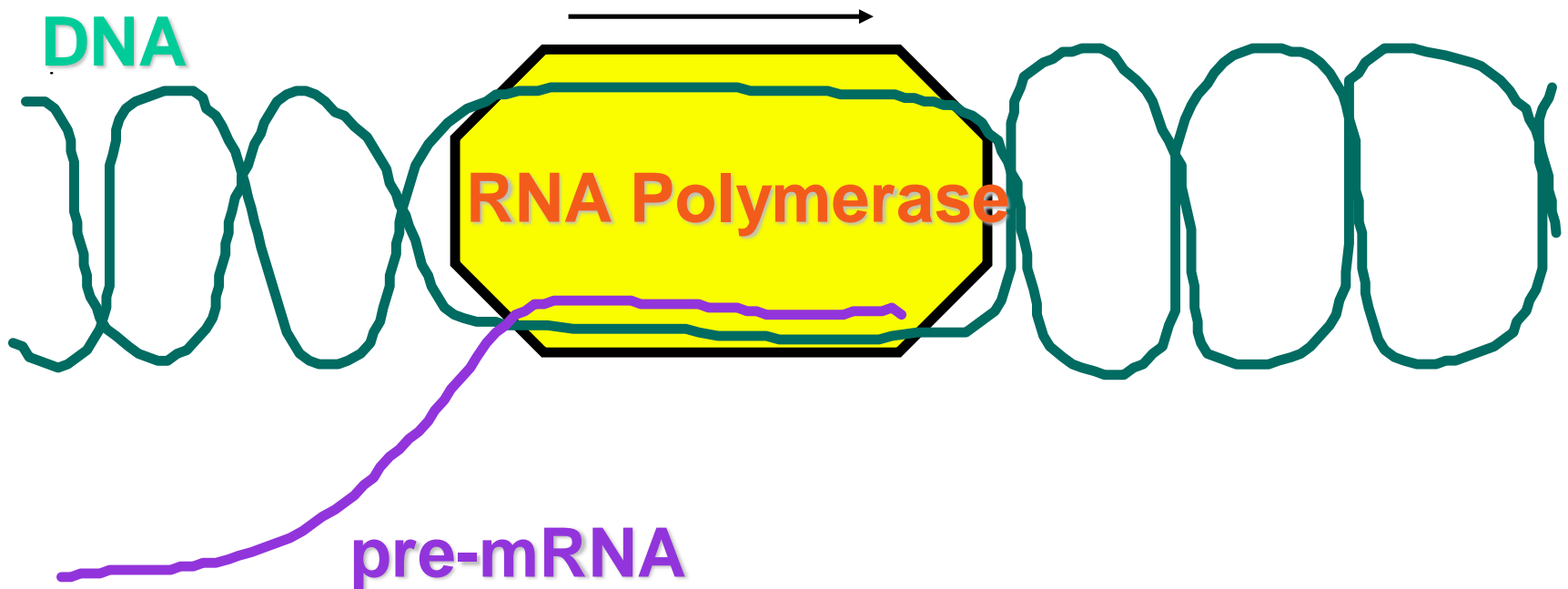
Overview of Transcription

- During **transcription** in the nucleus, a segment of DNA unwinds and unzips, and the DNA serves as a **template for mRNA formation**
- **RNA polymerase** joins the RNA nucleotides so that the **codons in mRNA are complementary** to the triplet code in DNA
- The transfer of information in the **nucleus** from a **DNA** molecule to an **RNA** molecule
- Only 1 **DNA** strand serves as the **template**
- Starts at promoter **DNA** (TATA box)
- Ends at terminator **DNA** (stop)
- When complete, **pre-RNA** molecule is released



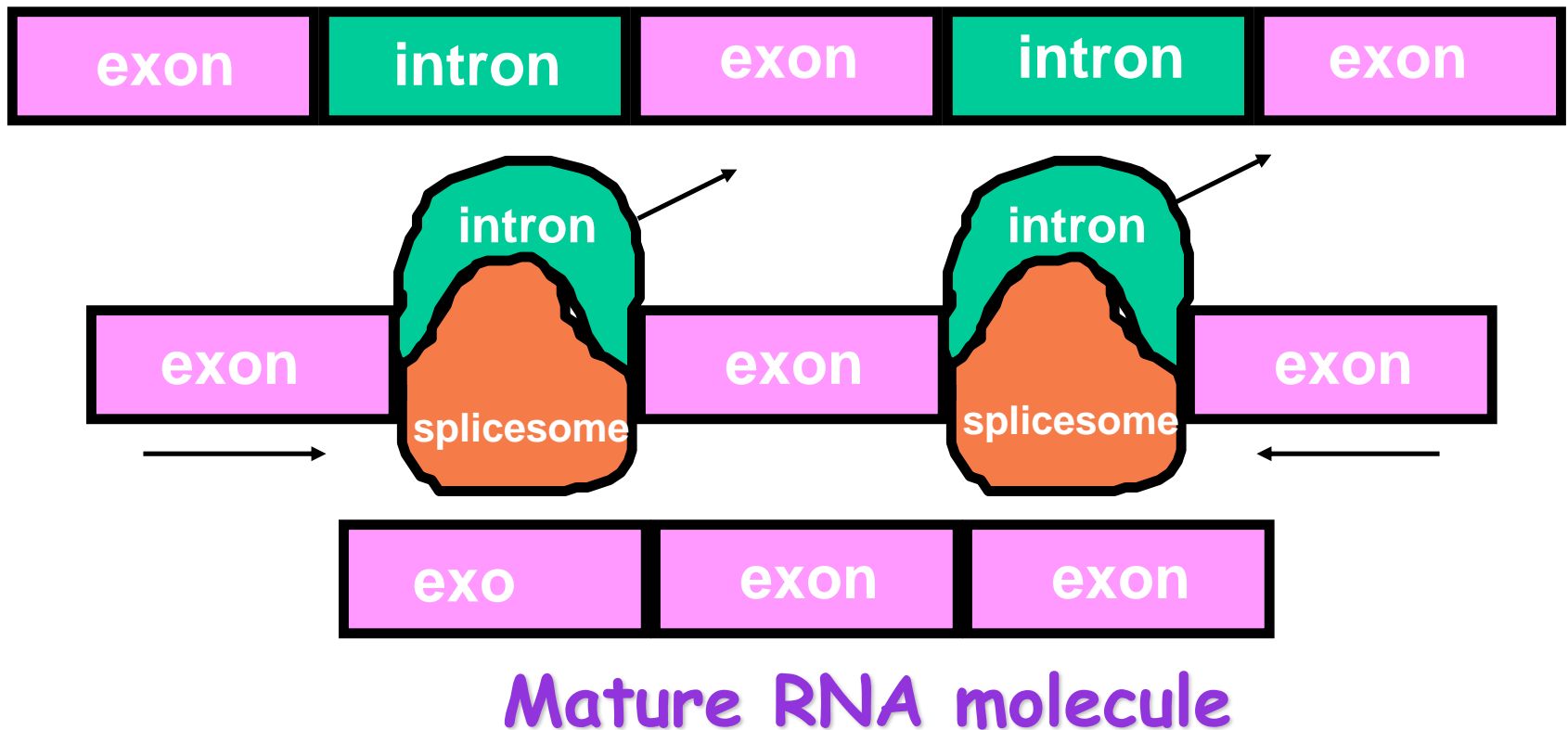
RNA Polymerase

- **Enzyme** found in the nucleus
- **Separates** the two DNA strands by **breaking the hydrogen bonds** between the bases
- Then moves along one of the DNA strands and **links RNA nucleotides** together



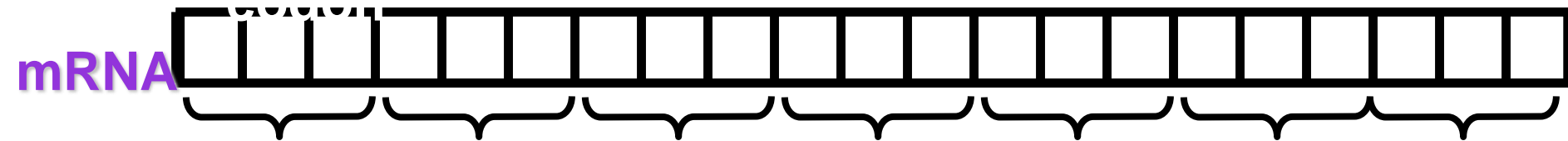
Processing Pre-mRNA

- Also occurs in the **nucleus**
- **Pre-mRNA** made up of segments called **introns & exons**
- **Exons code for proteins, while introns do NOT!**
- Introns spliced out by **splicesome-enzyme** and exons re-join
- End product is a **mature RNA** molecule that leaves the nucleus to the cytoplasm

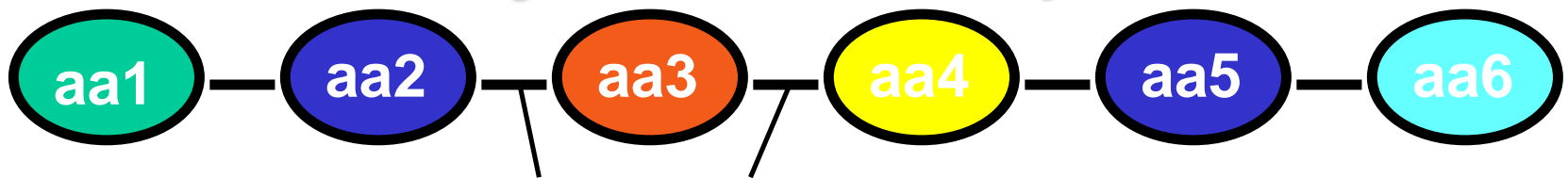


Messenger RNA (mRNA)

- Carries the information for a specific protein
- Made up of **500 to 1000** nucleotides long
- Sequence of 3 bases called **codon**
- **AUG** – methionine or **start** codon
- **UAA, UAG, or UGA** – **stop** codons



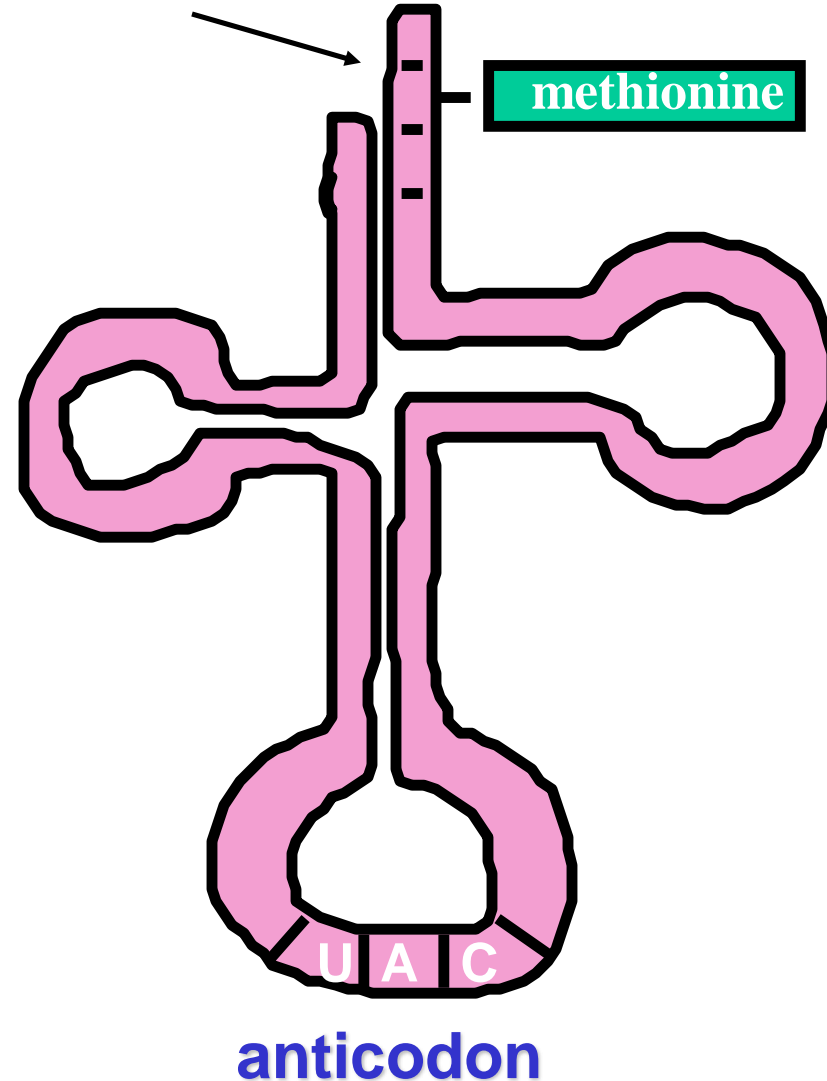
Primary structure of a protein



Transfer RNA (tRNA)

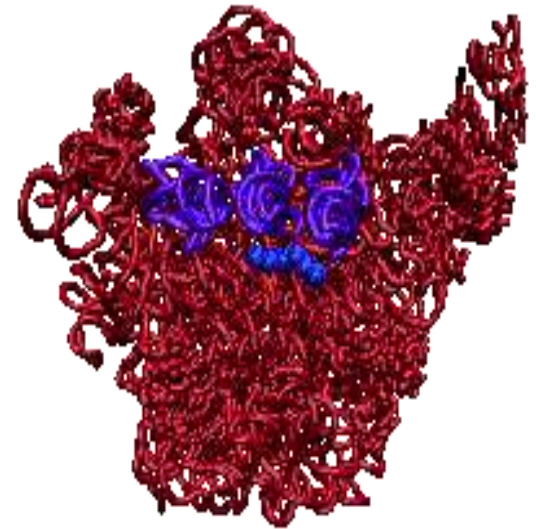
- Made up of **75 to 80 nucleotides** long
- Picks up the appropriate **amino acid** floating in the cytoplasm
- Transports **amino acids** to the **mRNA**
- Have **anticodons** that are complementary to **mRNA codons**
- Recognizes the appropriate **codons** on the **mRNA** and bonds to them with H-bonds
- **Four ATP's** are required for each amino acid added to the polypeptide chain: Two to "charge" the tRNA, one to carry the charged tRNA to the ribosome and one to move the ribosome to the next codon.

amino acid
attachment site



Ribosomal RNA (rRNA)

- Made up of rRNA is **100 to 3000 nucleotides** long
- Made inside the **nucleus** of a cell
- Associates with **proteins to form ribosomes**



Ribosomes

large and small

rRNA (40%)

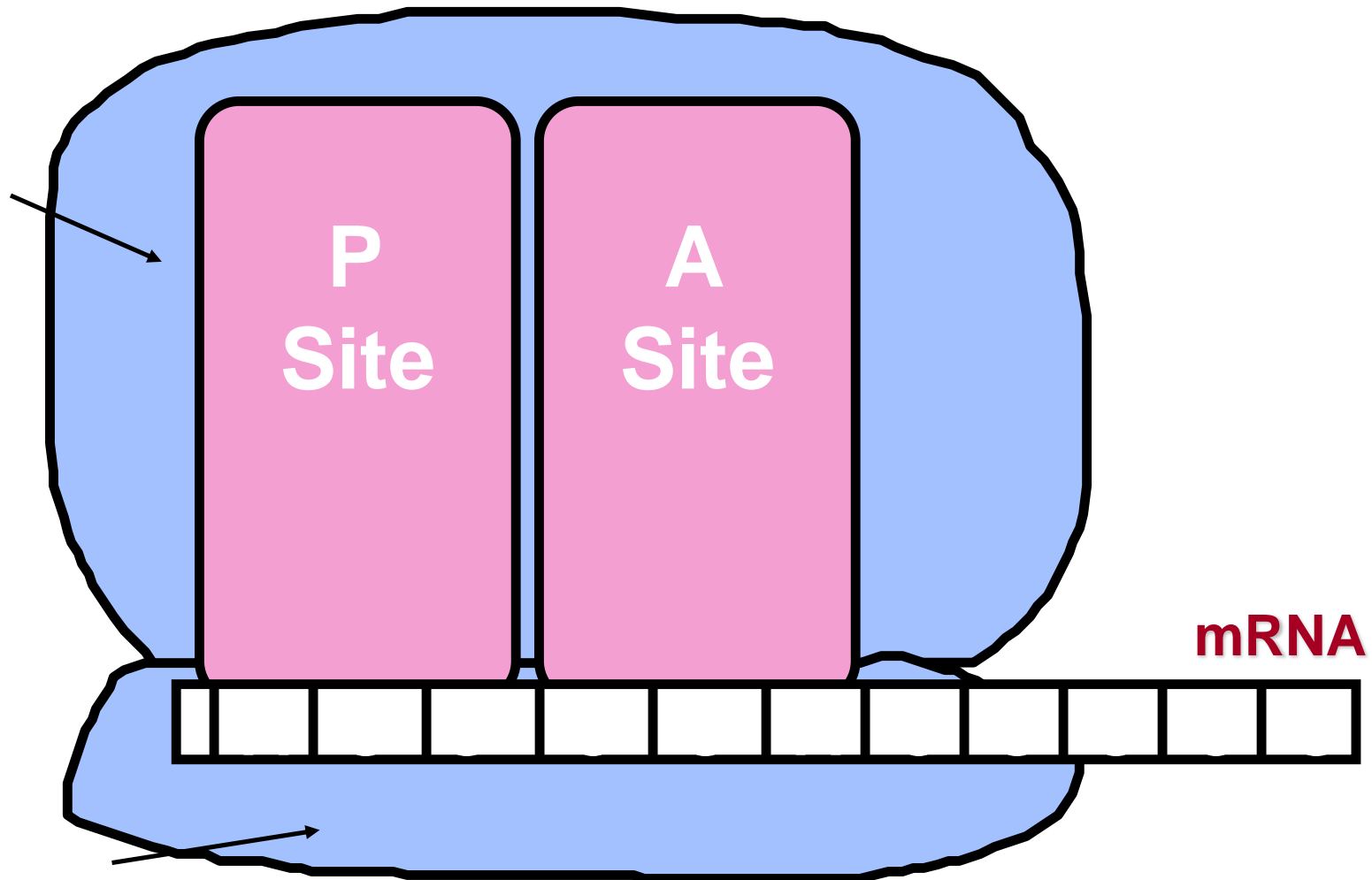
two sites

proteins (60%)

P and A

Ribosomes

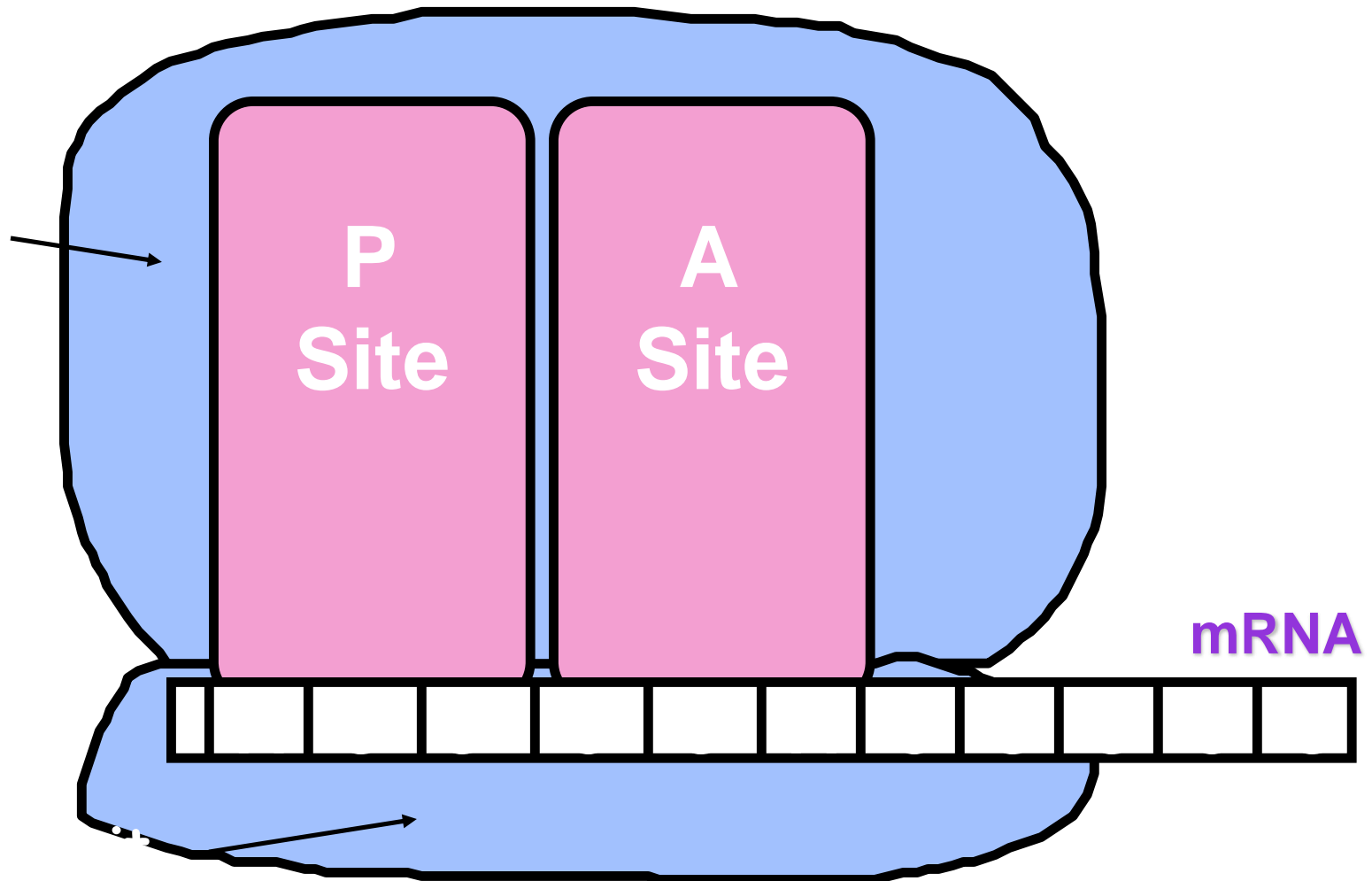
P= Peptide site
A= Amino acid site



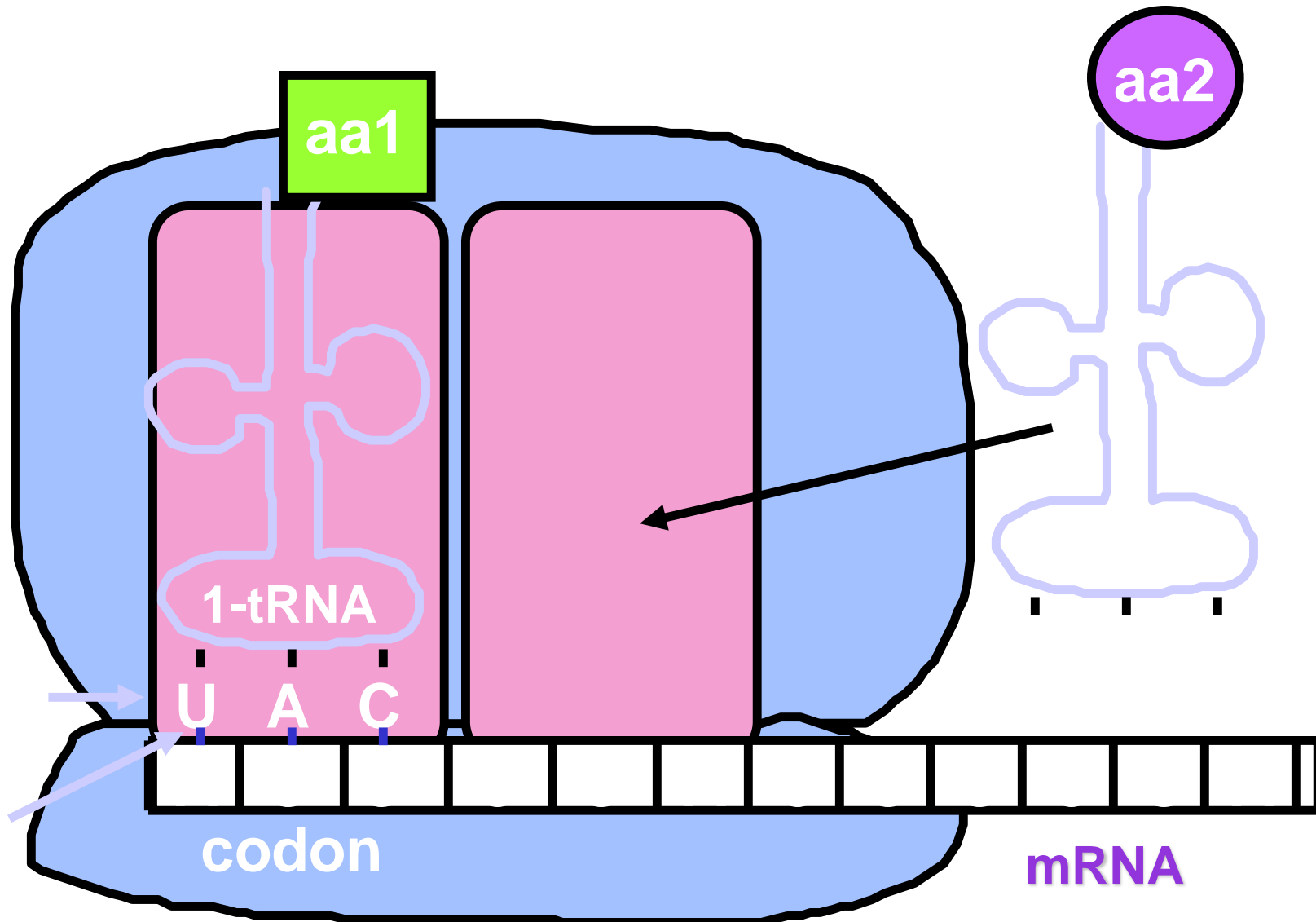
Translation

- **Synthesis of proteins** in the cytoplasm
- **Involves the following:**
 1. **mRNA** (codons)
 2. **tRNA** (anticodons)
 3. **ribosomes**
 4. **amino acids**
- **Three steps:**
 1. **initiation:** start codon (AUG)
 2. **elongation:** amino acids linked
 3. **termination:** stop codon (UAG, UAA, or UGA).

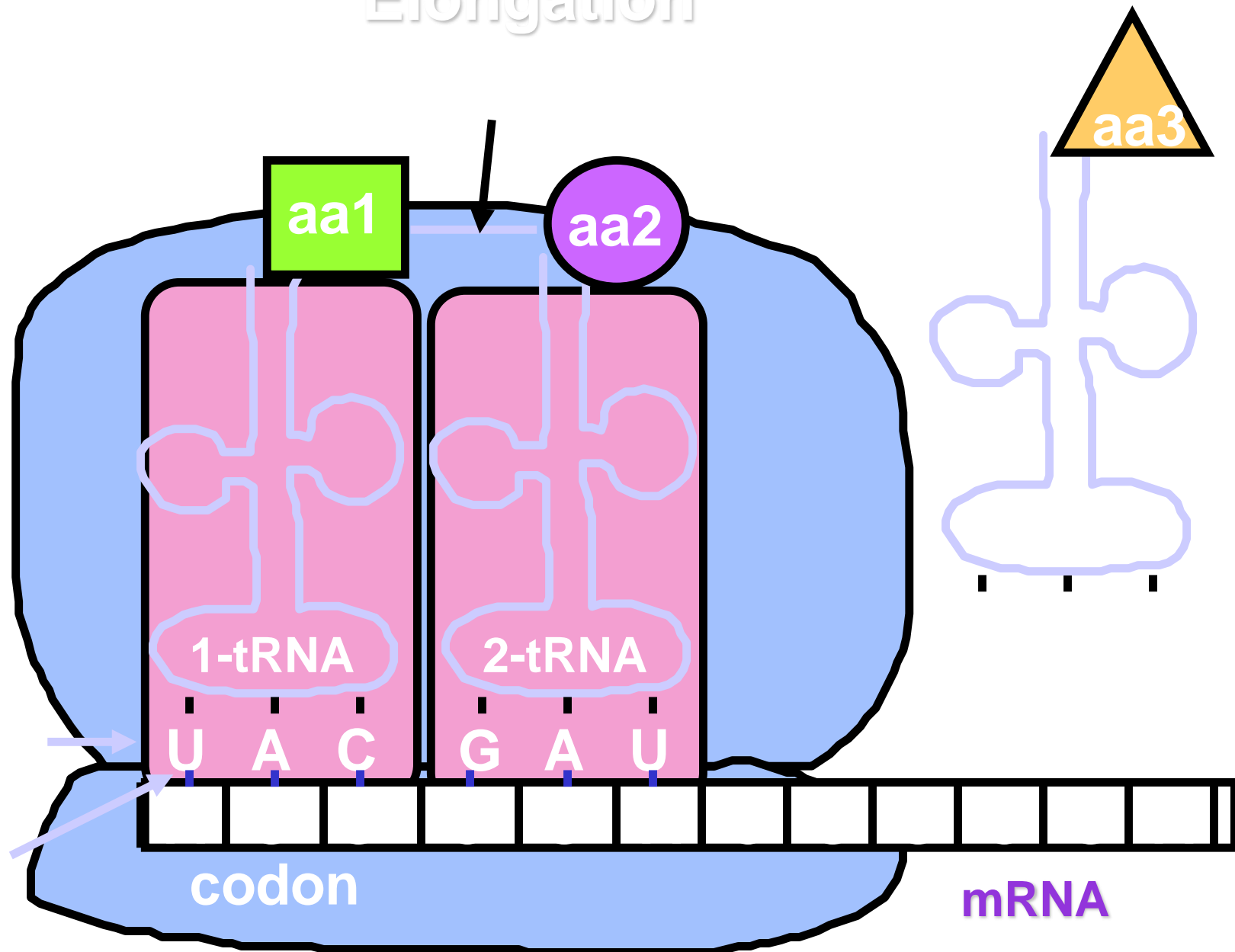
mRNA Codons Join the Ribosome

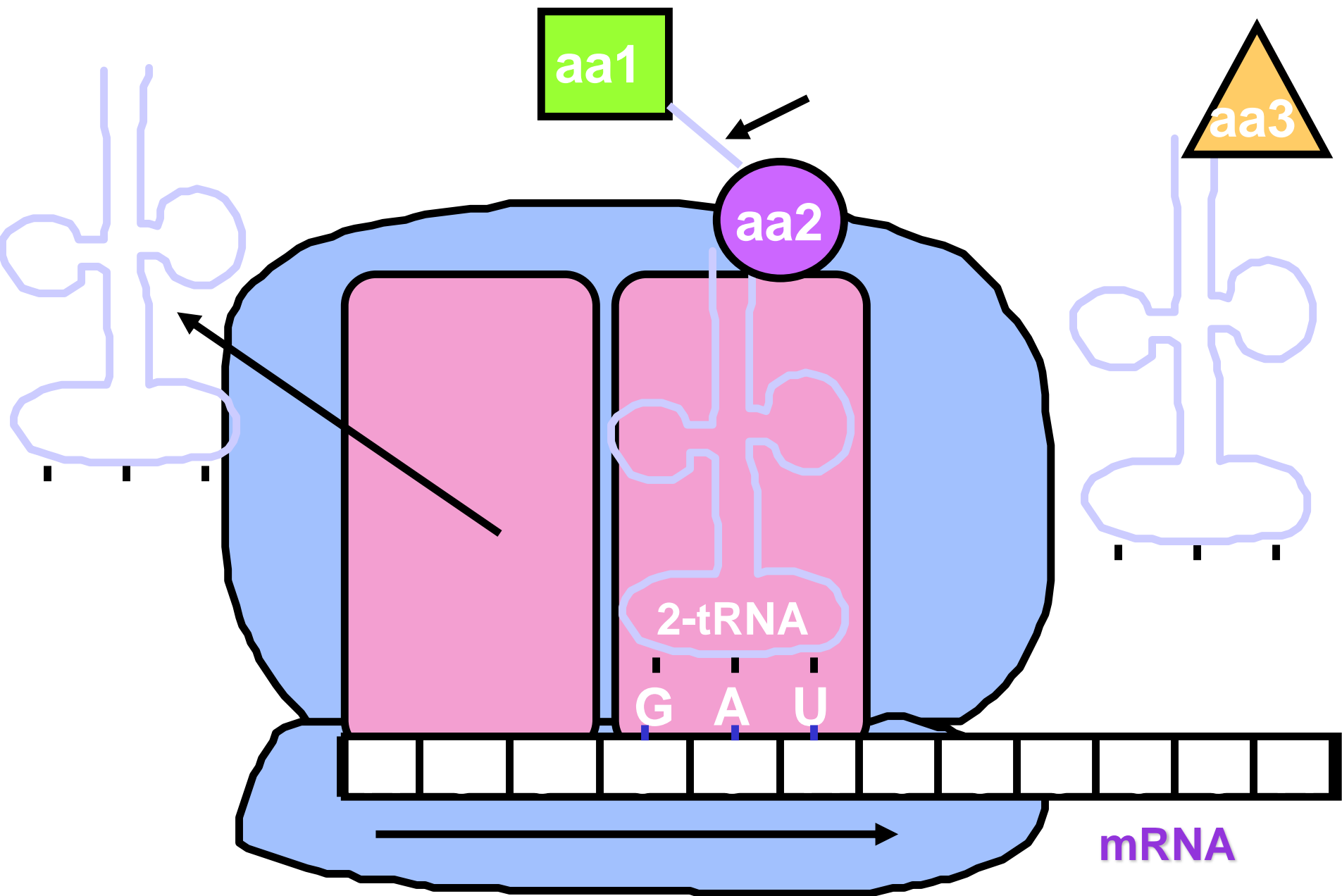


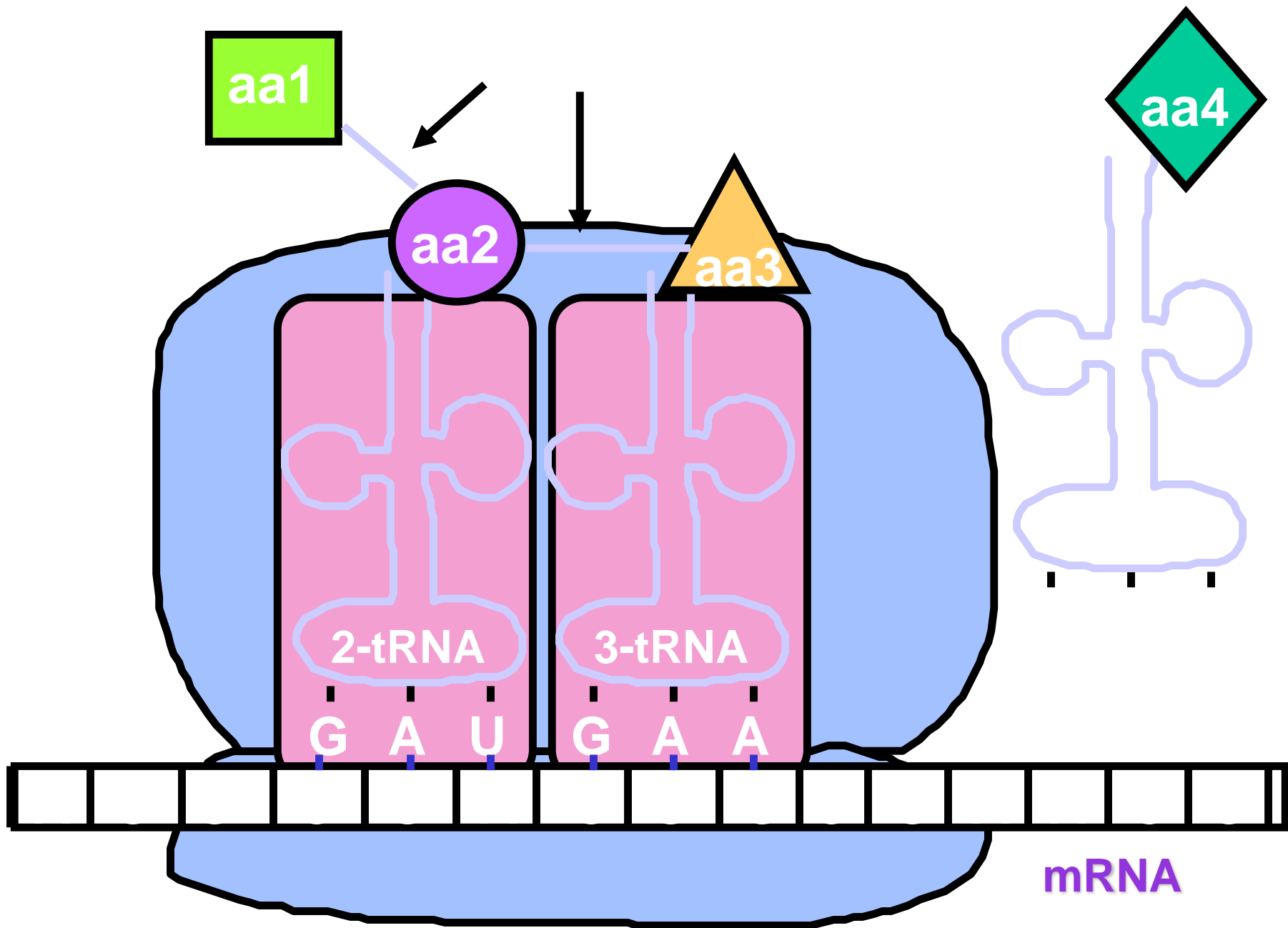
Initiation

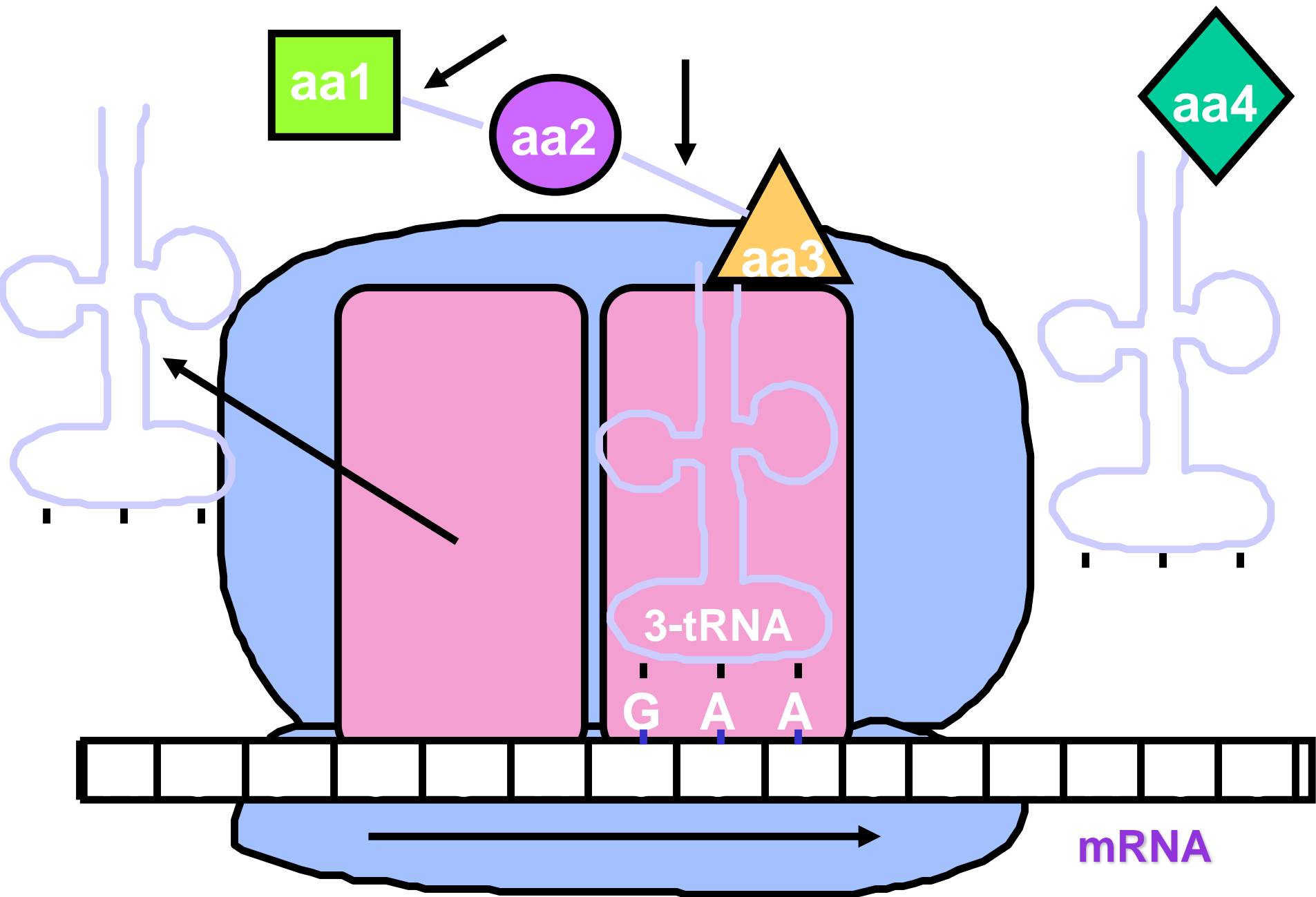


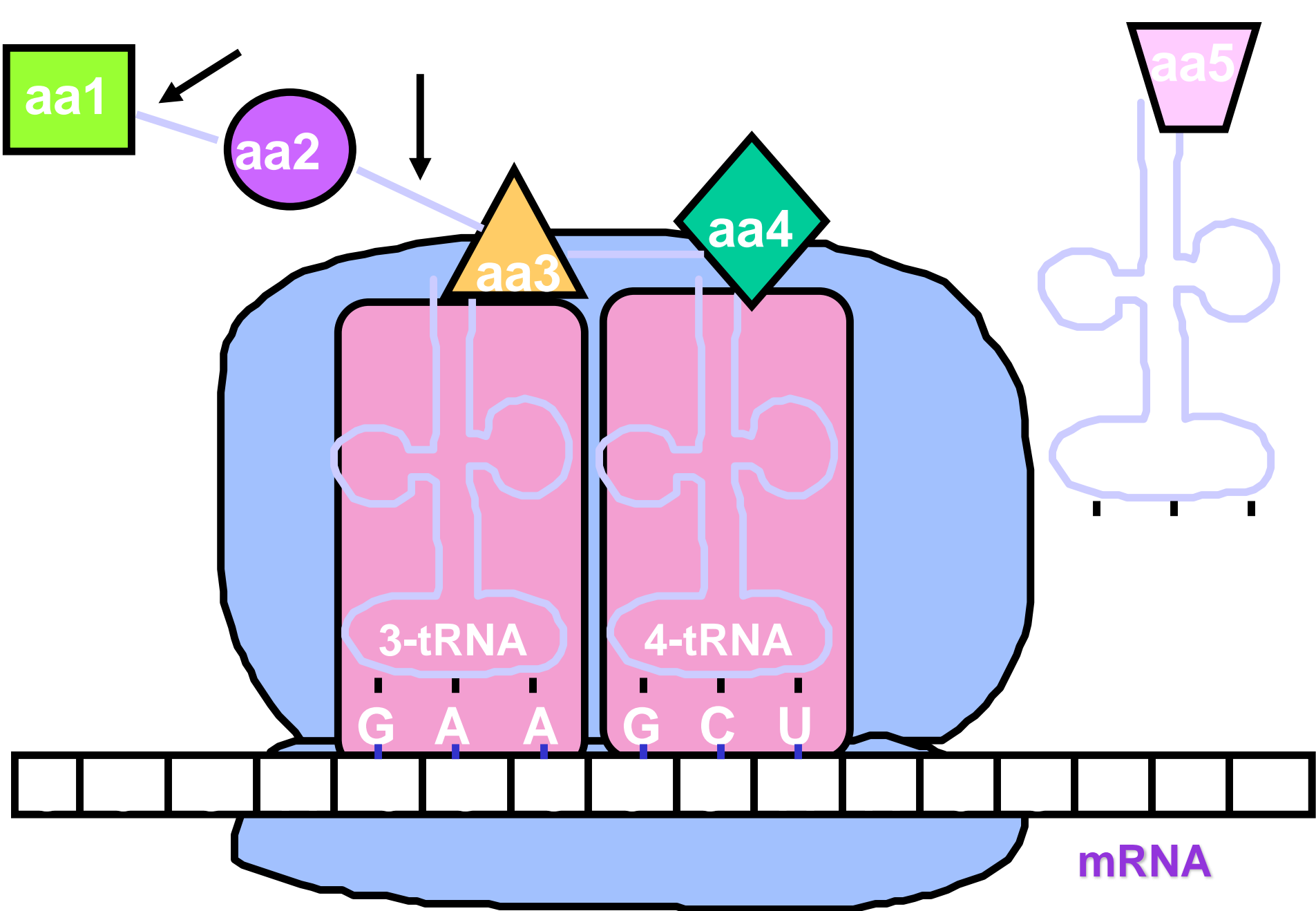
Elongation

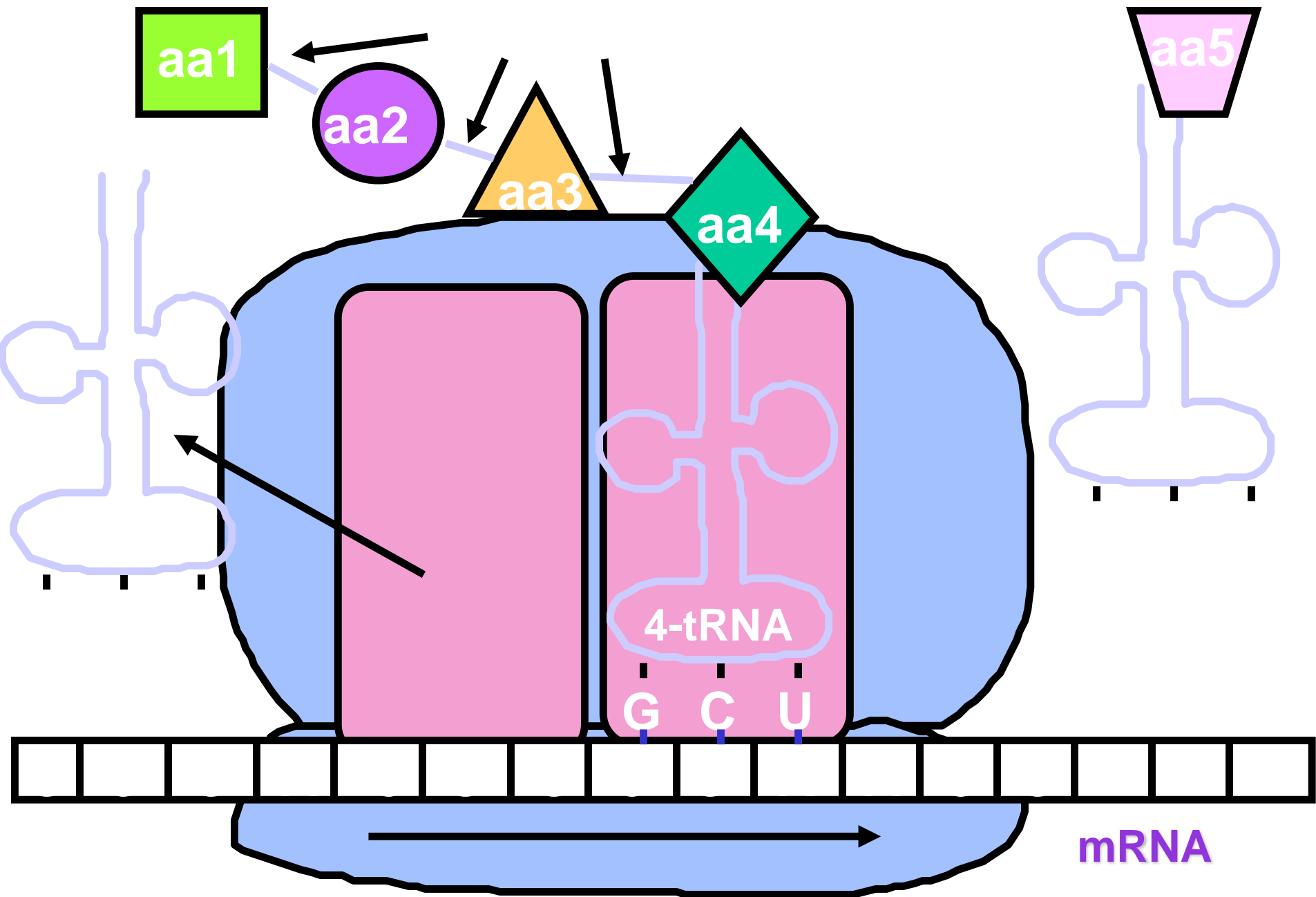


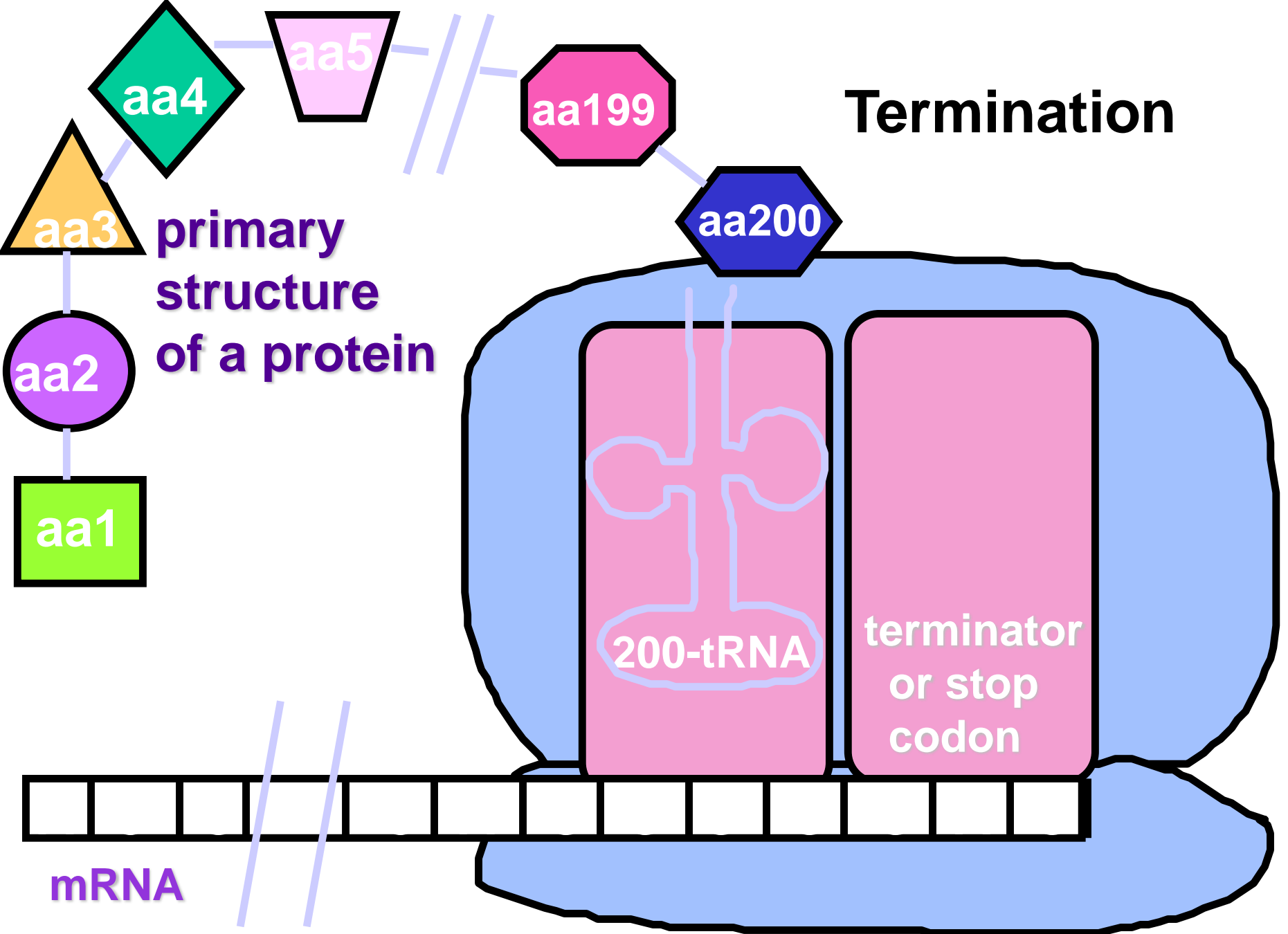






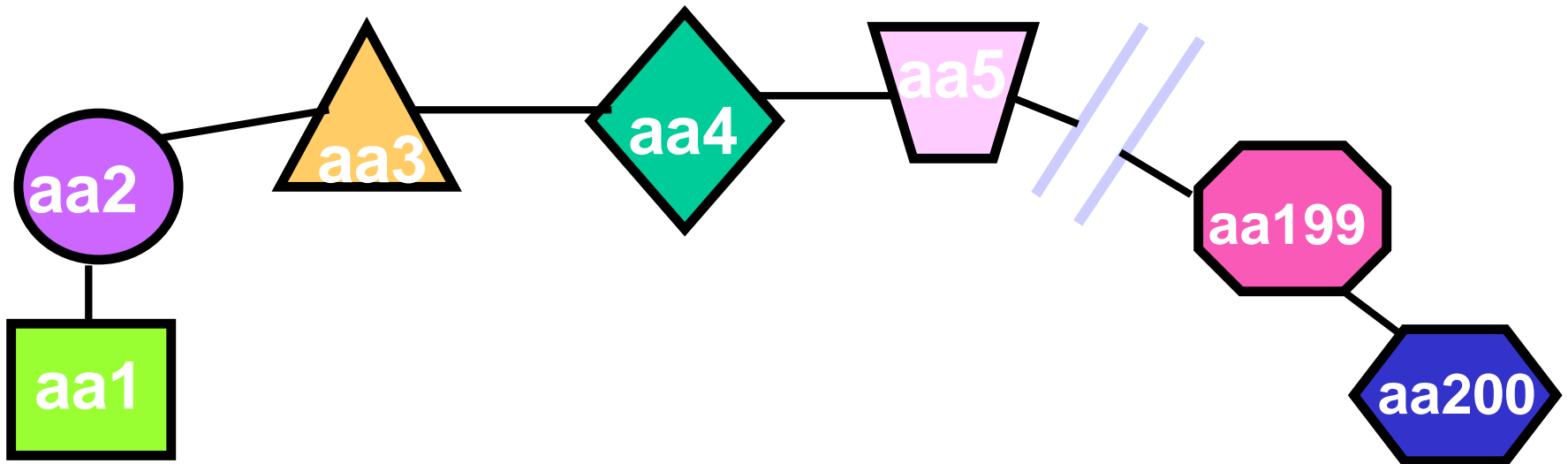






End Product –The Protein!

- The end products of protein synthesis is a **primary structure** of a protein
- A **sequence of amino acid** bonded together by peptide bonds



Classification of Proteins

Based on the chemical nature, structure, shape, and solubility, proteins are classified as:

- 1. Simple proteins:** They are composed of only amino acid residue. On hydrolysis, these proteins yield only constituent amino acids. It is further divided into:
Fibrous protein: Keratin, Elastin, Collagen
Globular protein: Albumin, Globulin, Glutelin, Histones
- 2. Conjugated proteins:** They are combined with non-protein moiety. Eg. Nucleoprotein, Phosphoprotein, Lipoprotein, Metalloprotein, etc.
- 3. Derived proteins:** They are derivatives or degraded products of simple and conjugated proteins. They may be
Primary derived protein: Proteans, Metaproteins, Coagulated proteins
Secondary derived proteins: Proteoses or albumoses, peptones, peptides.

Structure of a protein

1. Primary Structure

•The primary structure of a protein consists of the amino acid sequence along the polypeptide chain. Amino acids are joined by **peptide bonds**.

2. Secondary Structure

•**Alpha-helix**: The α -helix is a right-handed coiled strand. The side-chain substituents of the amino acid groups in an α -helix extend to the outside. Hydrogen bonds form between the oxygen of the C=O bond of an amino acid to hydrogen of the N-H group of the fourth amino acid below it in the helix.

•**Beta Sheet**: The hydrogen bonding in a β -sheet is between strands (inter-strand). The sheet conformation consists of pairs of strands lying side-by-side. The carbonyl oxygens in one strand hydrogen bond with the amino hydrogens of the adjacent strand.

3. Tertiary Structure

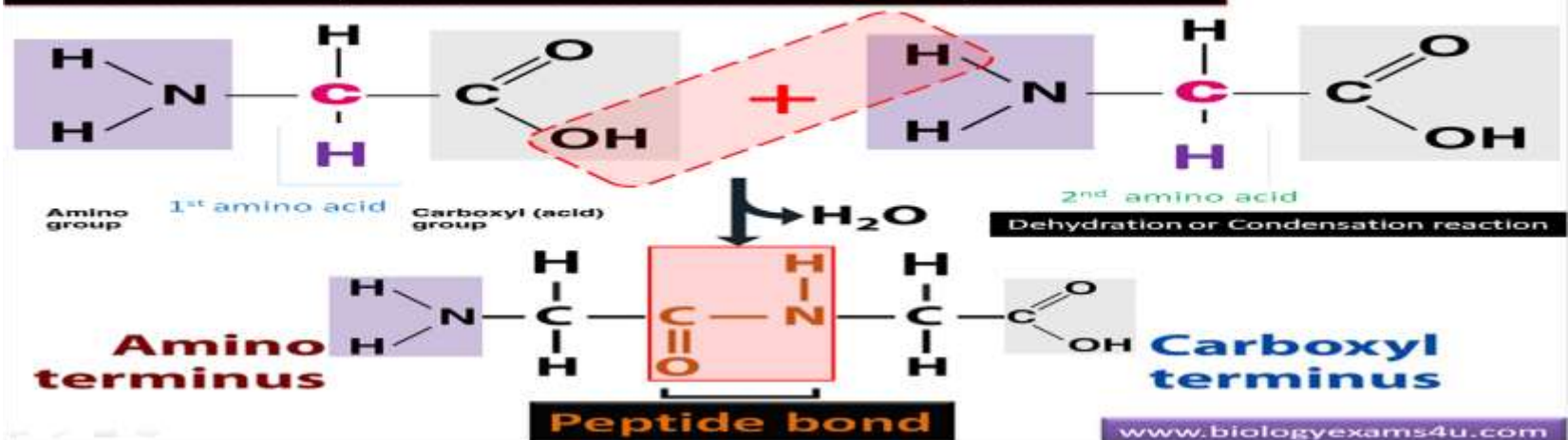
•The tertiary structure of a protein refers to its overall three-dimensional conformation.

•The types of interactions between amino acid residues that produce the three-dimensional shape of a protein include hydrophobic interactions, electrostatic interactions, and hydrogen bonds, all of which are non-covalent and Covalent disulfide bonds also occur.

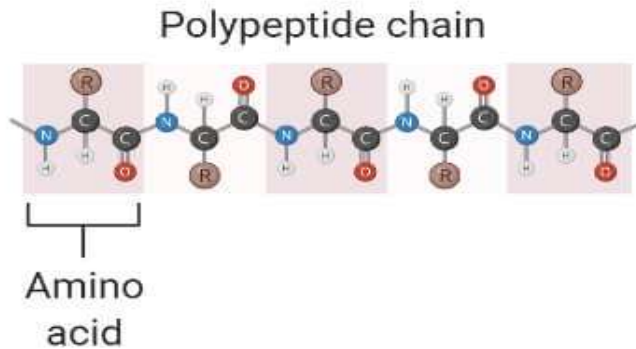
4. Quaternary Structure

•Quaternary structure refers to the interaction of one or more subunits to form a functional protein, using the same forces that stabilize the tertiary structure. It is the spatial arrangement of subunits in a protein that consists of more than one polypeptide chain.

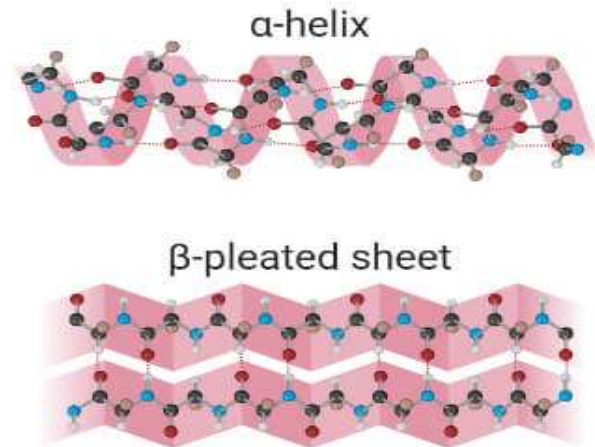
1. How is Peptide Bond formed in protein?



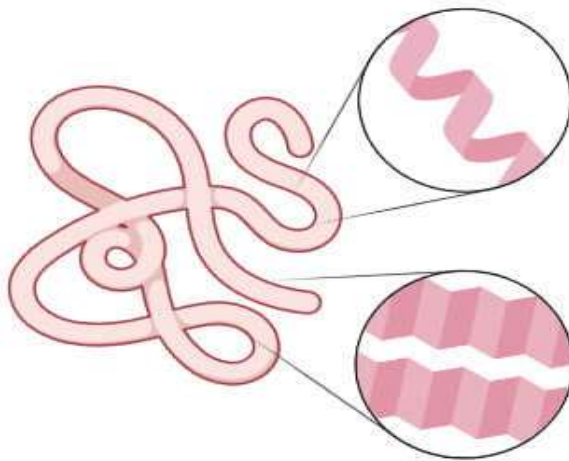
Primary structure



Secondary structure



Tertiary structure



Quaternary structure



Classes of Protein Structure

The function of a protein depends heavily on its final structure. Tertiary and quaternary proteins are both ***functional proteins*** with a 3D structure. However, the type of structure can vary significantly between different proteins.

There are two main classes of 3D protein structure: ***globular proteins*** and ***fibrous proteins***.

Globular Proteins

Globular proteins are usually round and ball-shaped. They usually have ***metabolic functions***, for example, they may be ***enzymes*** or ***antibodies***. Haemoglobin is an example of a globular protein.

Fibrous Proteins

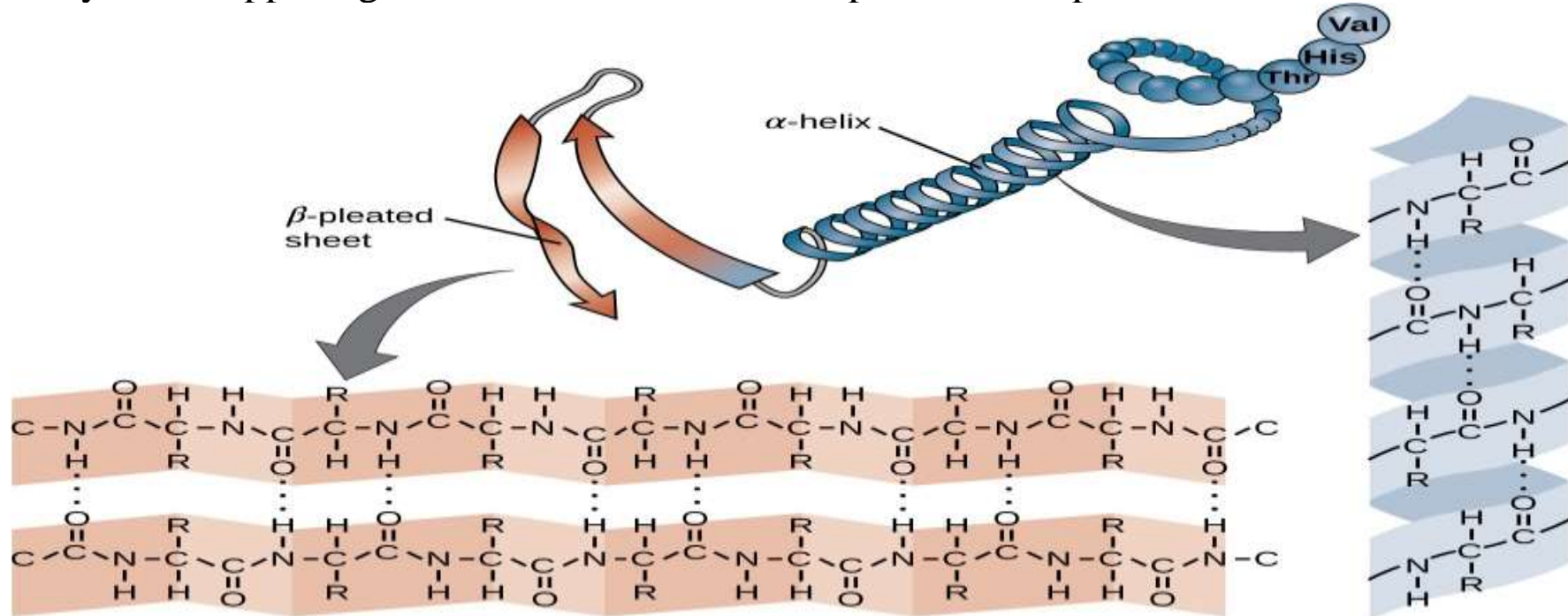
Fibrous proteins are long and narrow and usually have a ***structural function***. Examples of fibrous proteins include ***collagen*** (found in bones, muscle, and skin) and ***keratin*** (the material that makes up hair, nails, and feathers).

Functional Classification of Proteins

- Enzymes (biochemical catalysts)-In living organisms, almost all reactions are catalyzed by specific proteins called enzymes. They have a high catalytic power, increasing the rate of the reaction in which they are involved at least by factor 10^6 . Eg: Glucose isomerase
- Transport proteins -Many small molecules, organic and inorganic, are transported in the bloodstream and extracellular fluids, across the cell membranes, Eg: hemoglobin, that carries oxygen from the alveolar blood vessels to tissue capillaries;
- Storage proteins-Examples are: ferritin, that stores iron intracellularly in a non-toxic form
- Mechanical support-Proteins have a pivotal role in the stabilization of many structures. Examples are α -keratins, collagen and elastin.
- Movement-Example-the contraction of the muscle fibers (of which myosin is the main component);
- Nerve transmission.-An example is the receptor for acetylcholine at synapses.
- Hormones- regulatory functions Example is insulin,
- Protection against harmful agents.-The antibodies or immunoglobulins are glycoproteins that recognize antigens expressed on the surface of viruses, bacteria and other infectious agents.
- Storage of energy-Proteins in particular the amino acids that constitute them, act as energy storage, second in size only to the adipose tissue, that in particular conditions, such as prolonged fasting, may become essential for survival

Secondary structure of a protein

- The most common type of secondary structure in proteins is the α -helix.
- Fingernails and toenails are also made up of alpha-helices
- Beta Sheets form the core of globular proteins
- The end of a polypeptide chain can either be the N-terminus or the C-terminus. The N-terminus is the end that contains the free amino group, and the C-terminus is the end that contains the free carboxyl group.
- If two beta-strands run in the same direction, then it is a parallel beta-pleated sheet, and if they run in opposing directions, then it is an antiparallel beta-pleated sheet.



ALPHA HELIX VERSUS BETA PLEATED SHEET

Right-handed coiled
rod-like structure

Hydrogen bonds
form within the
polypeptide chain

Hydrogen bonds form
between N-H group of
one amino residue
with C=O group of
another amino acid

-R groups of the
amino acids are
oriented outside of
the helix

Alpha Helix can be
a single chain

Alpha Helix has only
one type

Alpha helix prefers Ala,
Leu, Met, Phe, Glu, Gln,
His, Lys, Arg amino
acids

Sheet-like
structure

Beta sheets are
formed by linking two
or more beta strands
by H bonds

Hydrogen bonds are
formed in between the
neighboring N-H and
C=O groups of adjacent
peptide chains.

-R groups are
directed to both
inside and outside
of the sheet

Beta Sheet cannot
exist as a single beta
strand

Beta Sheet can be
parallel, anti-parallel
or mixed

Beta sheet prefers Tyr,
Trp, (Phe, Met), Ile,
Val, Thr, Cys

Clinical Implications of Secondary Structure of Proteins



- The secondary structure of a protein can be altered by either a mutation in the primary sequence of amino acids that make up the protein or by extreme conditions that force the proteins to denature or lose their shape.
- Spongiform encephalopathy, and Amyloidosis are two classes of disease involving changes in the secondary structure of proteins.
- Both involve the misfolding of proteins into Beta sheets, and the presence of these proteins leading to tissue damage.

- Structural databases
- Protein visualizing tools
- Secondary structure prediction algorithms


Structural Databases

- Important in solving real problems in molecular biology
- Protein Databank - PDB Established in 1972 at Brookhaven National Laboratory (BNL)
- Sole international repository of macromolecular structure data
- Moved to Research Collaboratory for Structural Bioinformatics
<http://www.rcsb.org>

NDB: Nucleic Acid Structure Database



About NDBStandardsEducationToolsSoftwareDownload

**NUCLEIC ACID
DATABASE**

A Portal for Three-dimensional Structural Information about Nucleic Acids
As of 7-Sep-2022 number of released structures: 12269

Search DNASearch RNAAdvanced Search

Enter an NDB ID or PDB ID
Search for released structures

Welcome to the NDB

The NDB contains information about experimentally-determined nucleic acids and complex assemblies. Use the NDB to perform searches based on annotations relating to sequence, structure and function, and to download, analyze, and learn about nucleic acids.

New: Try out the [Nucleic Acid Knowledgebase \(NAKB\) server \(beta release\)](#)


NAKB is the planned successor to NDB, as described in this article: Berman, H.M.; Lawson, C.L.; Schneider, B. Developing Community Resources for Nucleic Acid Structures. Life 2022, 12, 540. <https://doi.org/10.3390/life12040540>

Search Structures

Search DNA
Search DNA and its complexes

Search RNA
Search for RNA structures in the NDB archive or in the Non-Redundant list

Advanced Search
Search for structures based on structural features, chemical features, binding modes, citation and experimental information



Featured Tools


RNA 3D Motif Atlas, a representative collection of RNA 3D internal and hairpin loop motifs

Non-redundant Lists of RNA-containing 3D structures

RNA Base Triple Atlas, a collection of motifs consisting of two RNA basepairs

WebFR3D, a webserver for symbolic and geometric searching of RNA 3D structures

R3D Align, an application for detailed nucleotide to nucleotide alignments of RNA 3D structures



- The NDB contains information about experimentally-determined nucleic acids and complex assemblies.
- Use the NDB to perform searches based on annotations relating to sequence, structure and function, and to download, analyze, and learn about nucleic acids.
- <http://ndbserver.rutgers.edu/>

PDB- Protein Structure database

- A protein database contains the information about 3D structure of proteins.
- The PDB files contain experimentally decided 3D structures of biological macromolecules.
- The structural information of a protein can be determined by X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy methods.
- PDB allows searching for information regarding the structure, sequence, function, visualize , download and to assess molecules.
- The PDB files also contains information of data collected, molecule name, primary and secondary structure, ligand, atomic coordinates, crystallographic structure factors, NMR experimental data etc..
- The data are submitted by scientists from all over the world. PDB is maintained by Worldwide Protein Data Bank. All data in PDB are accessible to public

- Each entry in the PDB is provided with a unique identification number called the PDB ID. It is a 4 letter identification number which consist of both alphanumeric characters.
- Without a proper tool, the PDB file will be read as a text file that lists each atom and its numerical coordinates in 3-D space.
- There are databases which contain data derived from PDB.

For example

- Structural Classification of Proteins (SCOP) that groups different protein structures,
- HSSP (Homology-Derived Secondary Structure of Proteins) for 3D-structure and 1D- sequence of the protein,
- CATH for protein structure classification according to their evolution etc.



195,093 Structures from the PDB

1,000,361 Computed Structure
Models (CSM)

3D Structures ?

Enter search term(s), Entry ID(s), or sequence

Include CSM ?

[Advanced Search](#) | [Browse Annotations](#)[Help](#)

NEW! Computed Structure Models (CSM)

[Learn more](#)

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:



Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive

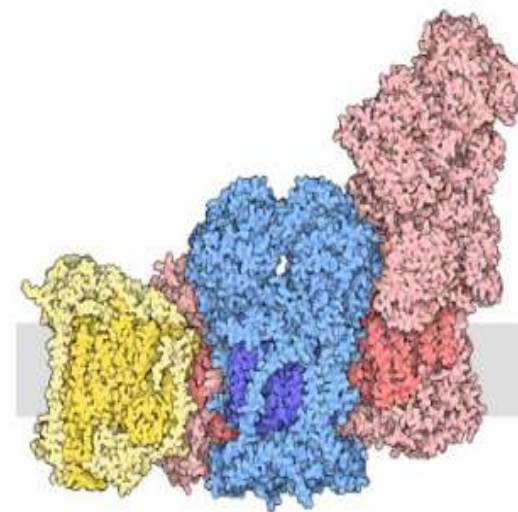


Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.



September Molecule of the Month



Respiratory Supercomplex

Protein visualizing tools

Uses

- To view the structures that are encoded by these atomic coordinate files (which have the extension .pdb), and
- to be able to manipulate the images to view the molecules from various perspectives, requires a molecular graphics visualization tool.

Protein visualizing tools

Rasmol/RasTop-

- RasMol is a molecular graphics program intended for the visualisation of proteins, nucleic acids and small molecules.
- The program reads in a molecule coordinate file and interactively displays the molecule on the screen in a variety of colour schemes and molecule representations.
- Currently available representations include depthcued wireframes, 'Dreiding' sticks, spacefilling (CPK) spheres, ball and stick, solid and strand biomolecular ribbons, atom labels and dot surfaces.



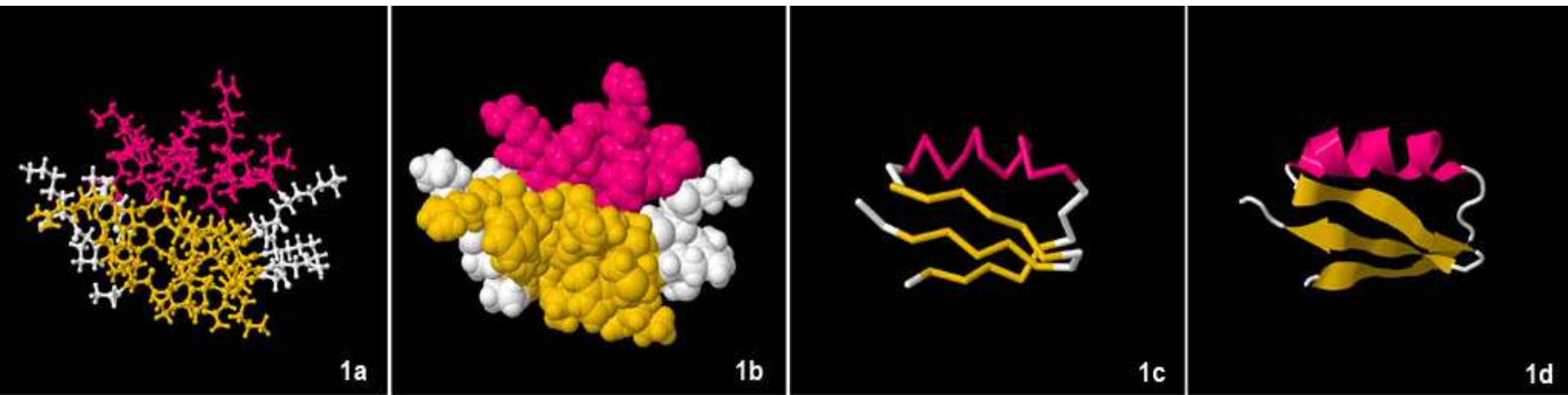
PyMol

- PyMOL is a molecular graphics system with an embedded Python interpreter designed for real-time visualization and rapid generation of high-quality molecular graphics images and animations.

- PyMol is a powerful molecule visualization software with the following main features:
- Able to produce high-quality graphics ready for publications.
- Able to create movies.
- Able to measure bond distances and angles.
- Has an extensive help system.
- Structures can be sliced, diced, and reassembled on the fly and written out to standard files.
- Both command line interface and graphical user interface are provided.
- Python API is provided to access all functionalities.

Online Visualizer

- **A PDB CODE (e.g. 1B8G)**
- Go to the Structure Home Page-
<https://www.ncbi.nlm.nih.gov/sites/entrez?db=Structure>
- Enter the PDB code in the search box and press the Go button.
- Click a structure image to access its record page
- Scroll to the [molecular graphic](#) section and click on the spin icon to load an interactive view of the structure within the web page.
Alternatively, click on the launch icon to open the advanced (full feature) version of iCn3D, NCBI's web-based 3D structure viewer, in a separate window.



- One protein, one true structure, eight ways to look at it using a molecular visualization tool
- Pink is for helix and yellow for sheet

- Ball and stick model - 1a show the main atoms in all of the protein's amino acids.
- Space filling model - In 1b you can see how much space each atom actually occupies. This is closest to the protein's actual shape.
- Backbone model - In 1c, all the details of the different amino acids have been removed and all that remains is a stick drawing showing where the protein's backbone goes.
- Ribbon (or cartoon) model - In 1d the individual amino acid structures have again been removed, but the flattened ribbon areas highlight places where the amino acids come together to form spirals (helices) or sheets. This makes it easier to visualize important secondary structures in the protein.

Secondary structure prediction algorithm

- The secondary structure prediction methods can be either
- ab initio based, which make use of single sequence information only, or
- homology based, which make use of multiple sequence alignment information.

- The *ab initio* methods, which belong to early generation methods, predict secondary structures based on statistical calculations of the residues of a single query sequence.
- The homology-based methods do not rely on statistics of residues of a single sequence, but on common secondary structural patterns conserved among multiple similar sequences.

Ab Initio–Based Methods

- It measures the relative propensity (Natural tendency) of each amino acid belonging to a certain secondary structure element.
- The propensity scores are derived from known crystal structures.
- Examples of ab initio prediction are the Chou–Fasman and Garnier, Osguthorpe, Robson (GOR) algorithms

The Chou–Fasman algorithm

- (<http://fasta.bioch.virginia.edu/fasta/chofas.htm>)
- determines the propensity or intrinsic tendency of each residue to be in the helix, strand, and β -turn conformation using observed frequencies found in protein crystal structures.

Homology-Based Methods

- This type of method combines the ab initio secondary structure prediction of individual sequences and alignment information from multiple similar sequences ($>35\%$ identity).
- The idea behind this approach is that close protein homologs should adopt the same secondary and tertiary structure.
- This homology based method has helped improve the prediction accuracy by another 10% over the second-generation methods.
- Eg.,- PHD

PHD – neural network algorithm for secondary structure prediction

- PHD are **neural network systems** (a sequence-to-structure level and a structure-structure level) to predict secondary structure (PHDsec)
- PHDsec focuses on predicting hydrogen bonds. The use of the evolutionary information held by a multiple sequence alignment increases the prediction accuracy.
- PHD uses two levels of Neural Networks • trained networks

PHD: Neural Networks

10 20 30 40 50 60 70

KTFTRCGLVHELKRGHGFEEENLMRNWVCLVEHESSRDTSKTNTNRNGSKDYGLFQINDRYWCSKGASPGKD

CCCCChHHHHHHHHHCCCCccccCCcEEEEEEeCCCCCCCCCCCCCCCCCCCCcEEEEECCCCCCCCCCCCCCCCC

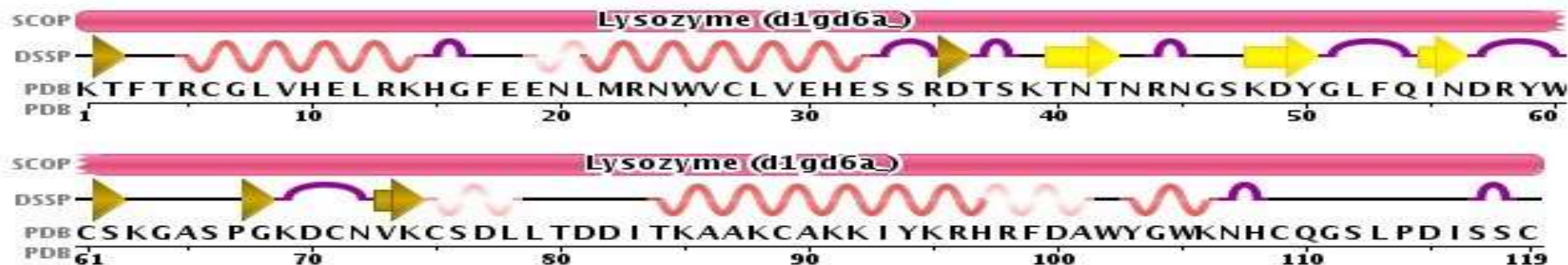
CNVKCSDLLTDDITKAAKCAKKIYKRHRFD AWYGWKNHCQGSLPDISSC

ceehHHHHhHHhChHHHHHHHHHHHHHHHCCCCcHHHHHHHHhCCCCCCCCC

CFSSP: Chou - Fasman

Query 1	KTFTRCGLVHELKRGHGFEEENLMRNWVCLVEHESSRDTSKTNTNRNGSKDYGLFQINDRYW	60
Helix 1	<----->	60
Sheet 1	EEEEEEEE	60
Turns 1	T	60
Query 61	CSKGASPGKDCNVKCSDLLTDDITKAAKCAKKIYKRHRFD AWYGWKNHCQGSLPDISSC	119
Helix 61	<----->	119
Sheet 61	EEEEEEEE	119
Turns 61	T T T	119

PDB 1GD6: STRUCTURE OF THE BOMBYX MORI LYSOZYME



GOR: GOR4

10 20 30 40 50 60 70

KTFTRCGLVHELKRGHGFEEENLMRNWVCLVEHESSRDTSKTNTNRNGSKDYGLFQINDRYWCSKGASPGKD

CCCCCCCCChhhhhCCCCccccccccccccccccCCCCCCCCCCCCCCCCCCCCccccccccccccCCCCC

CNVKCSDLLTDDITKAAKCAKKIYKRHRFD AWYGWKNHCQGSLPDISSC

ccceCCCCCCCCChhhhhhhhhhhhhhhhhCCCCcccccccccccccccc

How good are the methods?

- Single sequence, single residue methods
Chou & Fasman -50%
- Single sequence, multiple residues methods
GOR IV -65%
- Multiple sequence methods
PHD 71%

THANKYOU