Presentation slides

Machine Learning in Structural Biology: Interpreting 3D Protein Images

By

ALOYSIOUS KIKOMEKO – 2019/HD07/24856U

CHISOM SOREMEKUN - 2020/HD07/24246U

Introduction

- Proteins are essential molecules to all biological functions.
- All proteins are consisted of chains of 21 fundamental units knows as amino acids.
- And those fundamental units (amino acids) of proteins are connected by a chemical bond known as peptide bonds (type of a covalent bond).
- DNA molecules in cells encode information about how proteins is formed by combining amino acids (fundamental building blocks of all protein molecules).
- So cells are the factories that produce different types of proteins, which are essential for life.

Amino acids

Amino acids are the basic building blocks of proteins, the compounds that synthesize hormones and neurotransmitters in our bodies.

Amino acids help our bodies to grow, repair body tissue, maintain immunity and produce hormones that maintain body functions.

The 21 amino acids

The 21 amino acids different amino acids our bodies need are:

1 Alanine

2 Arginine

3 Asparagine

4 Aspartic acid

5 Cysteine

6 Glutamic acid

11 Leucine

12 Lysine

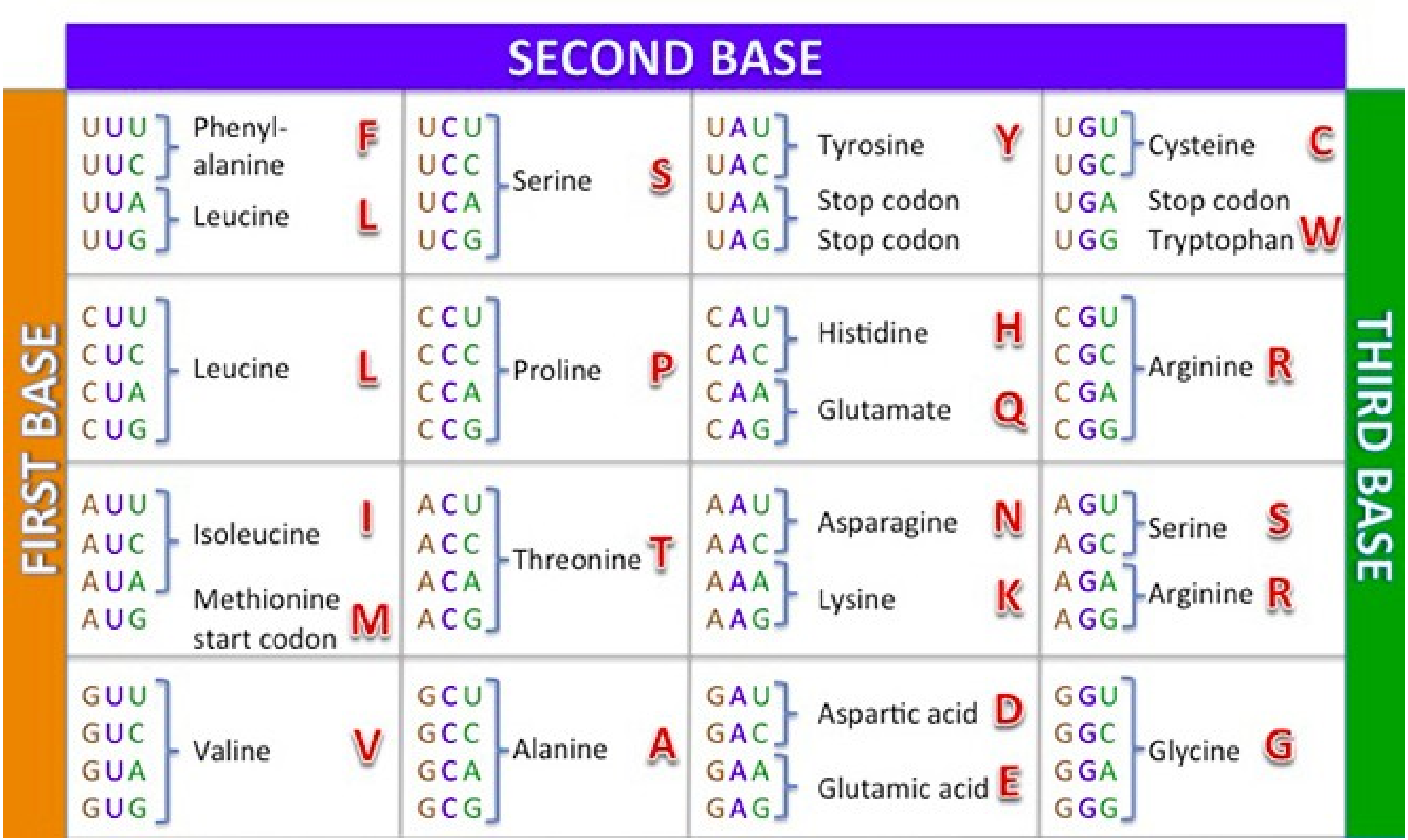
13 Methionine

14 Phenylalanine

15 Proline

16 Serine

The 21 amino acids codons



FIRST BASE	SECOND BASE				THIRD BASE
	UUU UUC Phenylalanine F UUA UUG Leucine L	UCU UCC UCA UCG Serine S	UAU UAC Tyrosine Y UAA Stop codon UAG Stop codon	UGU UGC Cysteine C UGA Stop codon UGG Tryptophan W	
	CUU CUC CUA CUG Leucine L	CCU CCC CCA CCG Proline P	CAU CAC Histidine H CAA CAG Glutamate Q	CGU CGC CGA CGG Arginine R	
	AUU AUC AUA Isoleucine I AUG Methionine start codon M	ACU ACC ACA ACG Threonine T	AAU AAC Asparagine N AAA AAG Lysine K	AGU AGC Serine S AGA AGG Arginine R	
	GUU GUC GUA GUG Valine V	GCU GCC GCA GCG Alanine A	GAU GAC Aspartic acid D GAA GAG Glutamic acid E	GGU GGC GGA GGG Glycine G	

https://sabal.uscb.edu/SIMMER_translation/

Protein structure

- Some examples of protein molecules in human body include hemoglobin in blood, various types of enzymes that help with chemical reactions, antibodies in immune system, and collagens in bone structure, etc.

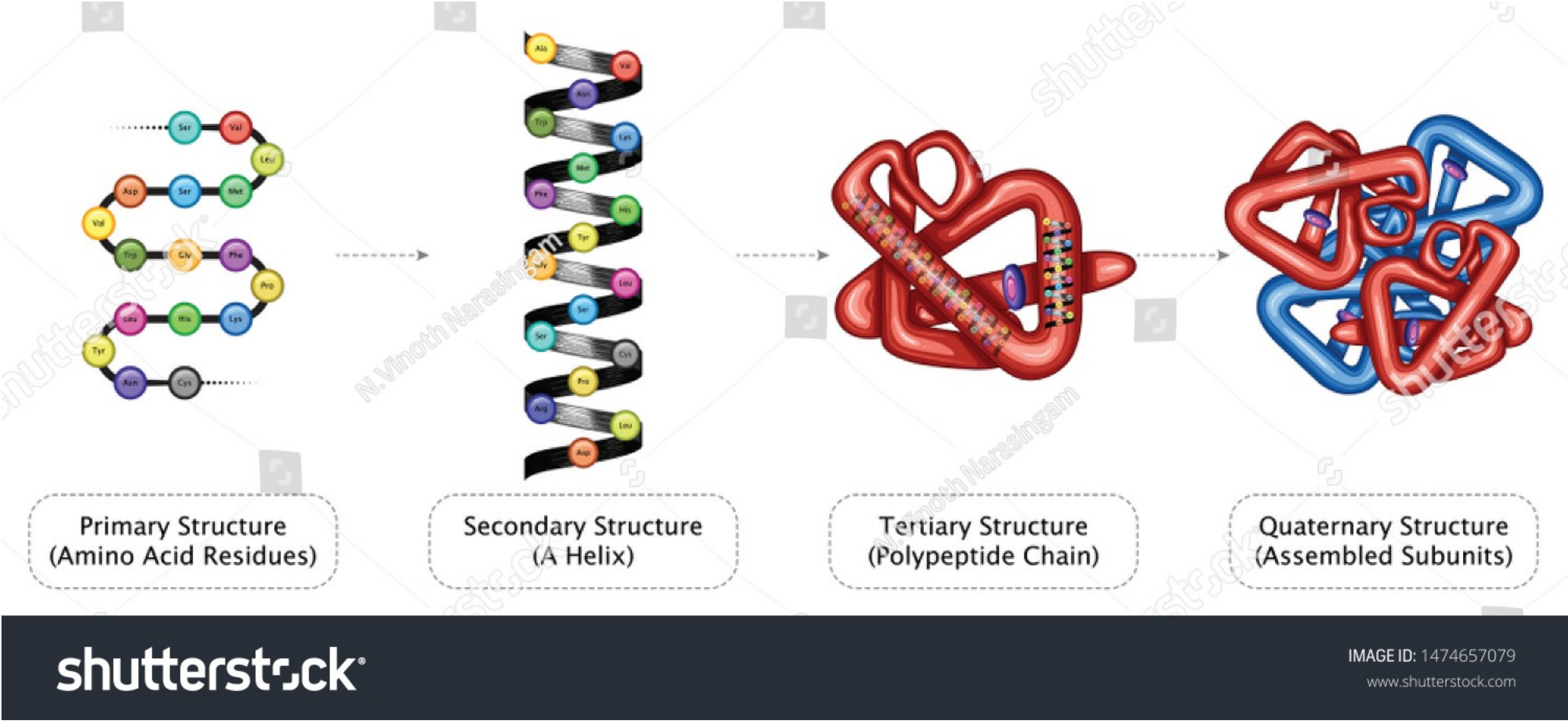
- Properties of different protein molecules are determined by its 3D structure. *So, determining 3D structure is very important and classic problem in biology*.

- Recent deep learning techniques are getting popular in protein structure detection problems.

Levels of protein structure

- In theory, there are 3 levels of protein structure representations.
- **Primary structure** - represented by the sequence of amino acids.
- **Secondary structure** - intermediate structure of proteins mainly consisted of major 2 types of secondary structure known as ***Alpha Helixes*** (sections of protein chain looks like spiral) and ***Beta Sheets*** (sections of protein chain looks like flat sheet).
- **Tertiary structure** - 3D structure of the protein.
- **Quaternary structure** - formed by combinations of several proteins

Levels of protein structure



Primary Structure
(Amino Acid Residues)

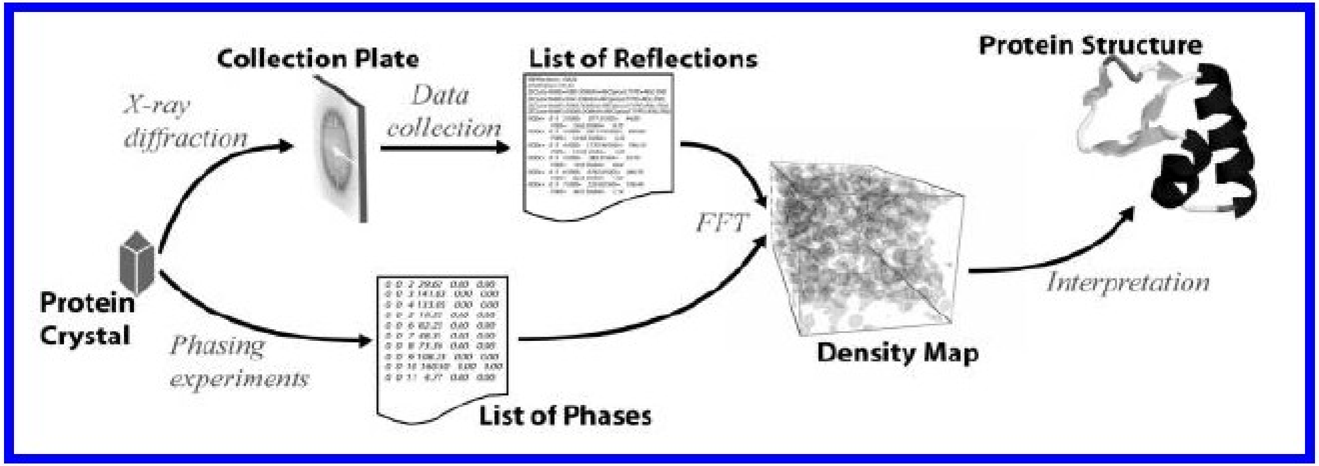
Secondary Structure
(A Helix)

Tertiary Structure
(Polypeptide Chain)

Quaternary Structure
(Assembled Subunits)

X-RAY CRYSTALLOGRAPHY

X-ray Crystallography is a scientific method used to determine the arrangement of atoms of a crystalline solid in three-dimensional space. Given a suitable target for structure determination, a crystallographer must produce and purify this protein in significant quantities. Only after estimating the phases can one compute the electron-density map - “density map”.



The quality of experimental maps as well as the sheer number of atoms in a protein makes interpretation difficult. Certain protein crystals produce a very narrow diffraction pattern, resulting in a poor-quality, “smoothed” density

Algorithmic background

Algorithms for automatically interpreting electron density maps draw heavily from the machine learning and statistics communities.

Machine learning Alogarithimic Models

1. Probabilistic models
2. Case-based reasoning
3. Neural networks

Approaches to automatic density map interpretation

A number of research groups have investigated automating the interpretation of electron-density maps.

ARP/WARP

The arp/warp (automated refinement procedure) software suite is a crystallographic tool for the interpretation and refinement of electron density maps. Arp/warp's warpntrace procedure was the first automatic interpretation tool successfully used for protein models.

RESOLVE

Resolve is a method developed by Terwilliger for automated model-building in poor-quality (around 3A) electron density maps

Approaches to automatic density map interpretation ….

- **TEXTAL**

Textal – another method for density map interpretation – was developed by Ioerger et al. Much like resolve, textal seeks to expand the limit of interpretable density maps to those with medium to low resolution (2 to 3A).

- **ACMI**

Acmi (automatic crystallographic map interpreter) is a recent method developed by DiMaio et al. for tracing protein backbones in poor-quality (3 to 4A resolution) density maps [19]. Acmi takes a probabilistic approach to electron density map interpretation, finding the most likely layout of the backboneunder some likelihood function.

References

[1] B. Rost and C. Sander (1993). Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol.

[2] G. Rhodes (2000). Crystallography Made Crystal Clear. Academic Press.

[3] J. Abrahams and R. De Graaff (1998). New developments in phase refinement. Curr Opin Struct Biol.

[4] S. Russell and P. Norvig (1995). Artificial Intelligence: A Modern Approach. Prentice Hall.

[5] T. Mitchell (1997). Machine Learning. McGraw-Hill.

[6] V. Lamzin and K. Wilson (1993). Automated refinement of protein models. Acta Cryst.