

# Course content

## Topics (lectures):

1. **Introduction** (1)
2. **Structure** (2-5)
3. **Single molecule mechanics** (6-9)
4. **Collective/emergent properties** (9-11)
5. **Student presentations** (11-13)

## Course structure:

1. **Introduction to topic**
2. **Awardees (1-2 per week)**
  - **History, first-person, second-person accounts (C)**
  - **Article, analysis of scientific work (E)**
3. **Discussion of topic, outlook**

# Recap of last time + loose ends

Philosophy of science and scientific breakthroughs:

- Kuhn
- Alternatives to Kuhn

What information do we need to decide if something is a breakthrough?

- context
- details about the work (facts)
- impact
  - citations on a topic before and after
  - scientific landscape before and after

What makes a breakthrough?

- nonlinear change in understanding / paradigm shift
- fills a knowledge gap
  - which was unfilled for a long time
  - better than other attempts by some-fold
- opens a new direction or create a new field. creative.

# Lecture 2: Introduction to structure

Today's goal: Biomolecules, history of structure

first half

- The stuff of life
- What is a structure
- Proteins as random walks + HP models

PBOC Chapter 1.1, 1.2, 8.1, 8.4

*last week*

second half

- History of structure

<https://knowablemagazine.org/article/living-world/2022/structural-biology-how-proteins-got-their-closeup>

Your mission:

Integrate information about the context of biomolecular structure.

- Open questions / challenges / goals of the field.

Name one open question in the field of biomolecular structure

Understand / explain the key developments in the field

- Human-sized vs. consortium-sized vs. field-sized breakthroughs

Give an example of a breakthrough described, and identify its "scale"

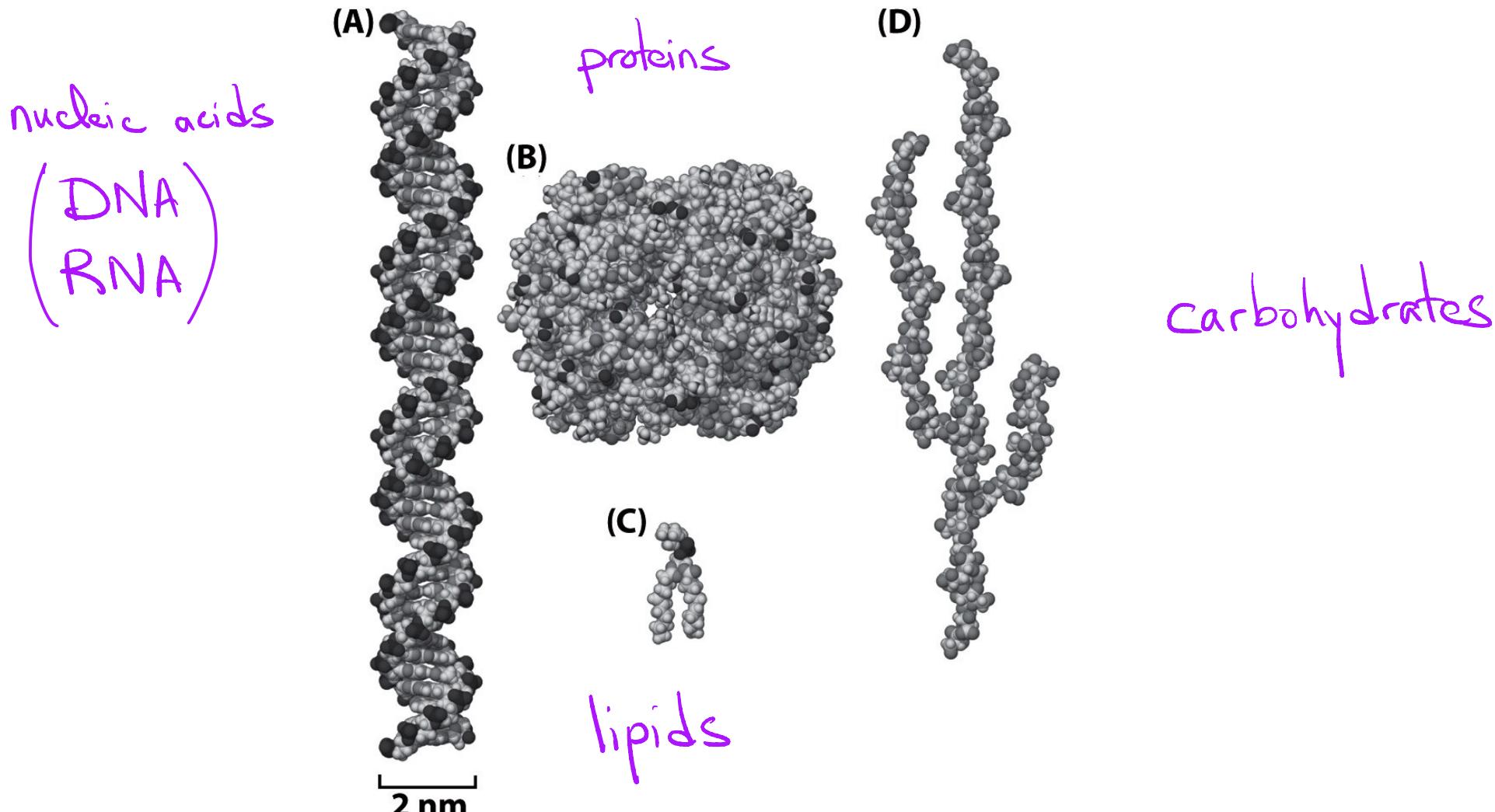
Classify "genre" of information sources

- expository: neutral, logical, balanced, objective
- argumentative: establishes an opinion, presents arguments to support as well as reasonable alternatives
- narrative: historical, tells a story or describes an experience

Identify the genre used in the media we listened to / watched

# The stuff of life

What macromolecules are cells made of?



# The stuff of life

&lt;divNucleic acids and proteins are polymer languages

&lt;divCENTRAL DOGMA of MOLECULAR

# BIOLOGY

## PROTEINS

NUCLEIC ACIDS

## ALPHABET

The diagram shows four nucleotides arranged in a square pattern. The top-left nucleotide is 'G', the top-right is 'T', the bottom-left is 'C', and the bottom-right is 'A'. Below this square, the word 'nucleotides' is written in a bold, black, sans-serif font.

## codon

## WORDS

The diagram shows a DNA segment represented by a horizontal line with two grey circles at the ends. Along this line, there are four vertical oval shapes, each containing a letter: 'G', 'A', 'C', and 'G' from left to right. This represents a sequence of four nucleotides in a DNA strand.

## SENTENCES

DNA



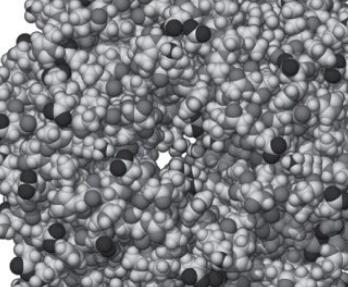
RNA polymerase mRNA

## transcription

## ALPHABET

## WORDS

## SENTENCES

A 3D space-filling model of a protein-DNA complex. The protein is shown in grey, and the DNA is represented by black spheres. The complex is a large, roughly spherical structure with a dense core of grey atoms and black spheres interspersed throughout.

ribosome | translation

## amino acids

The diagram illustrates the structure of a protein segment. On the left, an  $\alpha$  helix is shown as a spiral ribbon with amino acid residues labeled: R (Arginine), Y (Tyrrosine), E (Glutamate), L (Leucine), N (Asparagine), I (Isoleucine), and S (Serine) from top to bottom. On the right, a  $\beta$  strand is depicted as a vertical arrow with residues M (Met), T (Threonine), Y (Tyrrosine), and K (Lysine) attached to its side chain.

A 3D-style illustration of a coiled metal spring. The spring is light gray with a dark gray center. It is shown in a partially compressed state, with one end bent upwards and the other end bent downwards, creating a curved shape.

# Why structure?

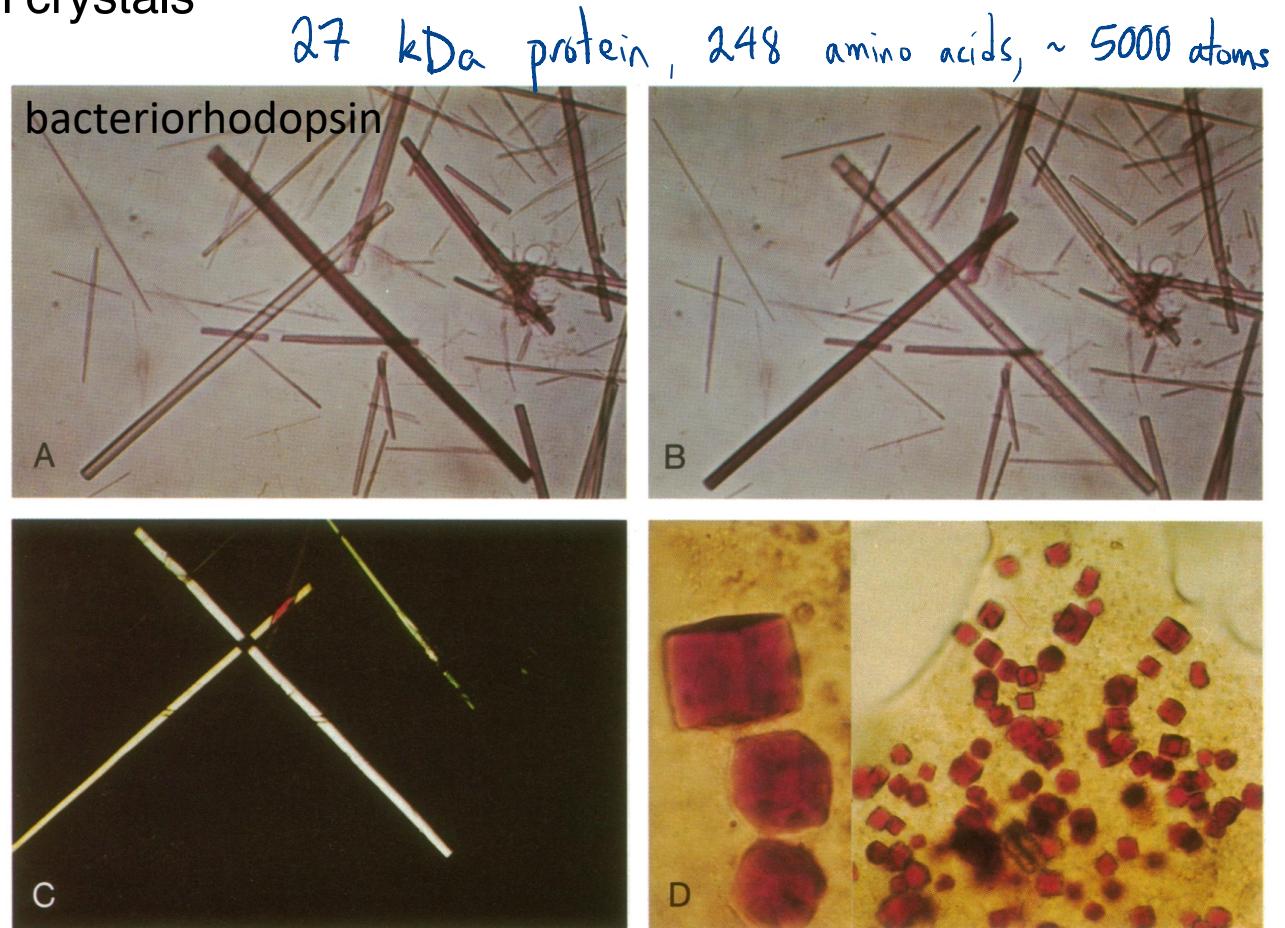
## Structure-function relationship

- Example of hemoglobin.
- Understand
  - fundamental curiosity
  - engineer new functions
  - treat disease

# What is a structure?

## Deterministic vs. statistical descriptions

Crystal: regular packing of atoms into unit cells  
Some proteins will form crystals



# What is a structure?

## Protein data bank: repository of structures



The screenshot shows the RCSB PDB website. At the top, there is a search bar with the placeholder "Enter search term(s), Entry ID(s), or sequence". Below the search bar are buttons for "Include CSM" and a magnifying glass icon. To the right of the search bar are "Advanced Search" and "Browse Annotations" links, and a "Help" link. The top navigation bar includes links for "PDB-101", "www.PDB", "EMDataResource", "Nucleic Acid Database", "wwPDB Foundation", and "PDB-Dev". Below the top bar is a navigation menu with the following items:

- Documentation
- General Help
  - Organization of 3D Structures in the Protein Data Bank
- Identifiers in PDB
- Computed Structure Models and RCSB.org
- Assessing the Quality of 3D Structures
- Data From External Resources Integrated Into RCSB PDB
- Symmetry Resources in the PDB
- Ligand Structure Quality in PDB Structures
- Web Services Overview
- X-ray Electron Density Maps
- Structures Without Legacy PDB Format Files
- Deposition Resources
- Software Supporters
- Membrane Protein Resources
- Website FAQ
- Glossary
- Search and Browse
- Exploring a 3D Structure

The "General Help" section is currently selected. Under "General Help", the "Organization of 3D Structures in the Protein Data Bank" link is also selected. The main content area displays the following text and links:

## Organization of 3D Structures in the Protein Data Bank

- Overview
  - Definitions
  - Relevance in Exploring the PDB
- Example

Video: [Entry, Entity, Assembly, and Instance](#)

## Overview

Biomolecules are hierarchical structures. For example, proteins are composed of linear chains of amino acids that (often) fold into compact subunits which then can associate into higher level assemblies with other proteins, small molecule ligands, and water or other solvent molecules. Biomolecules in the Protein Data Bank (PDB) archive are organized and represented using this hierarchy to simplify searching and exploration.

## Definitions

Four levels of hierarchy are commonly used: Entry, Entity, Instance, and Assembly:

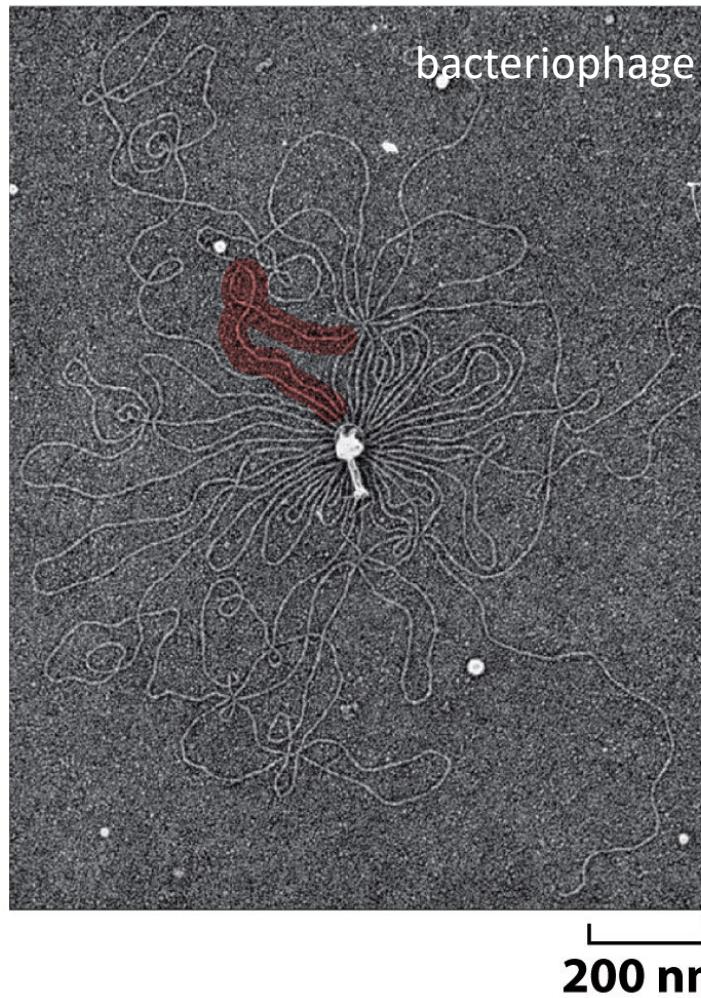
- An ENTRY is all data pertaining to a particular structure deposited in the PDB and is designated with a 4-character alphanumeric identifier called the PDB identifier or PDB ID (e.g., 2hbs).
- An ENTITY is a chemically unique molecule that may be polymeric, such as a protein chain or a DNA strand, or non-polymeric, such as a soluble ligand. Some entries may even have branched polymeric entities, such as oligosaccharides.
- An INSTANCE is a particular occurrence of an ENTITY. An ENTRY may contain multiple INSTANCES of an ENTITY, for example, many copies of a protein chain in a homooligomeric protein.
- An ASSEMBLY is a biologically relevant group of one or more INSTANCES of one or more ENTITIES that are associated with each other to form a stable complex and/or perform a function.

# What is a structure?

Deterministic vs. **statistical** descriptions

PB&C 8.1

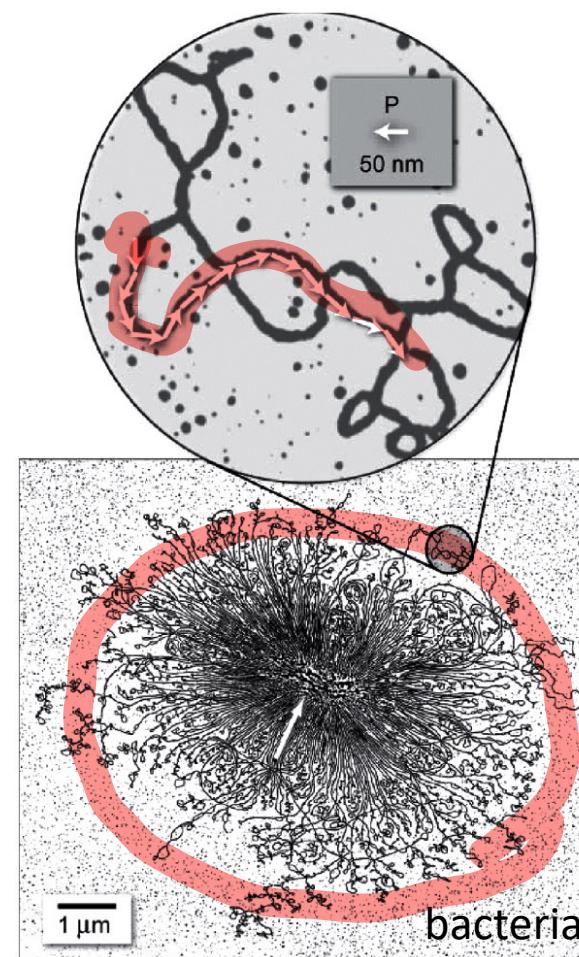
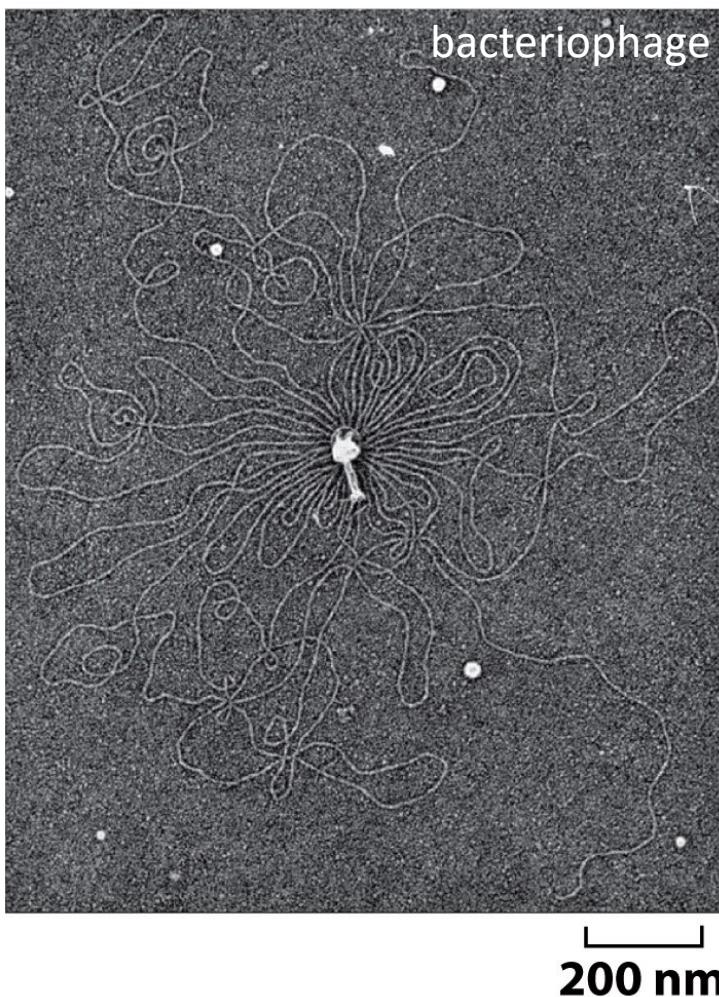
Average size and shape, fluctuations are important



# What is a structure?

Deterministic vs. **statistical** descriptions

Average size and shape, fluctuations are important



What can we learn  
about the molecule?

elasticity

$$r \propto m^n$$



"proportional to"

$$n = \frac{1}{2} \quad \text{or} \quad \frac{3}{5}$$

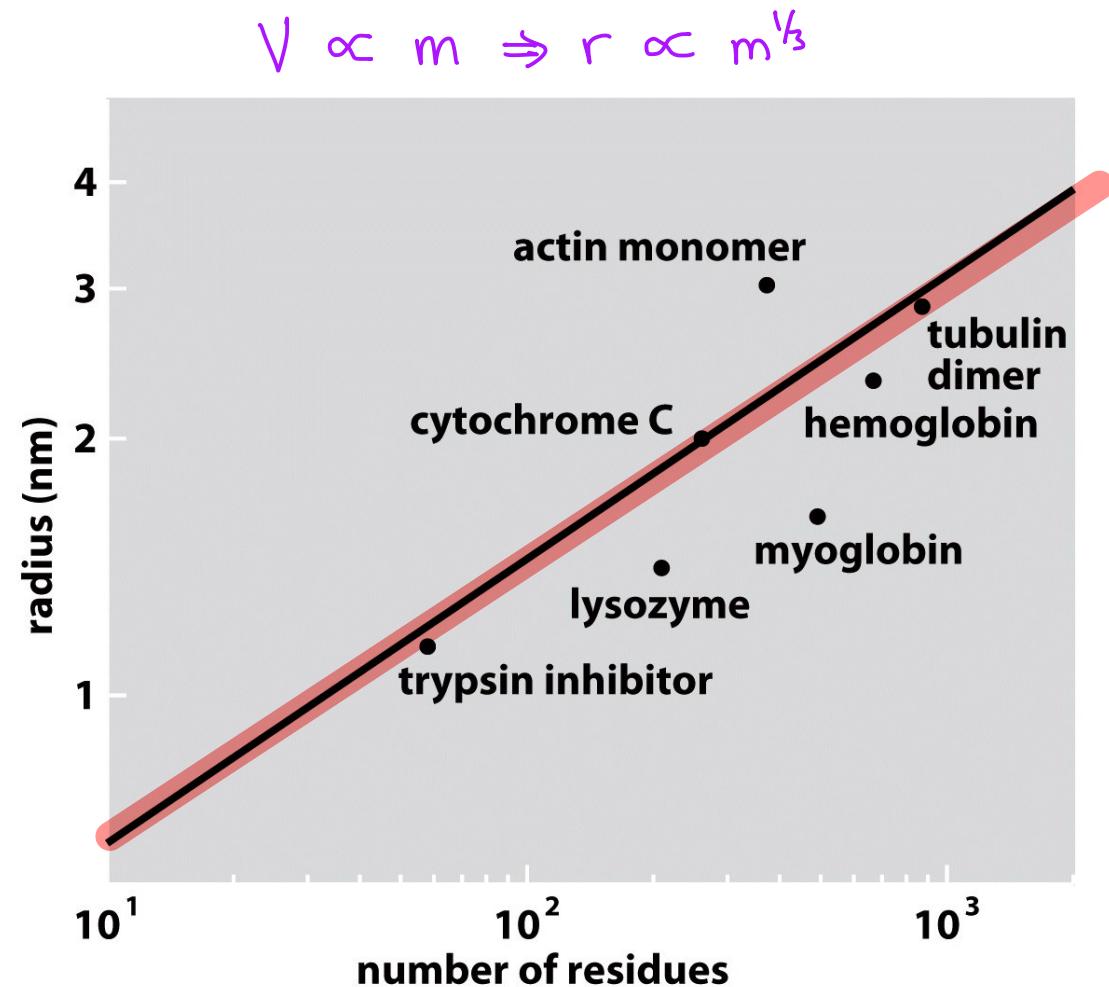
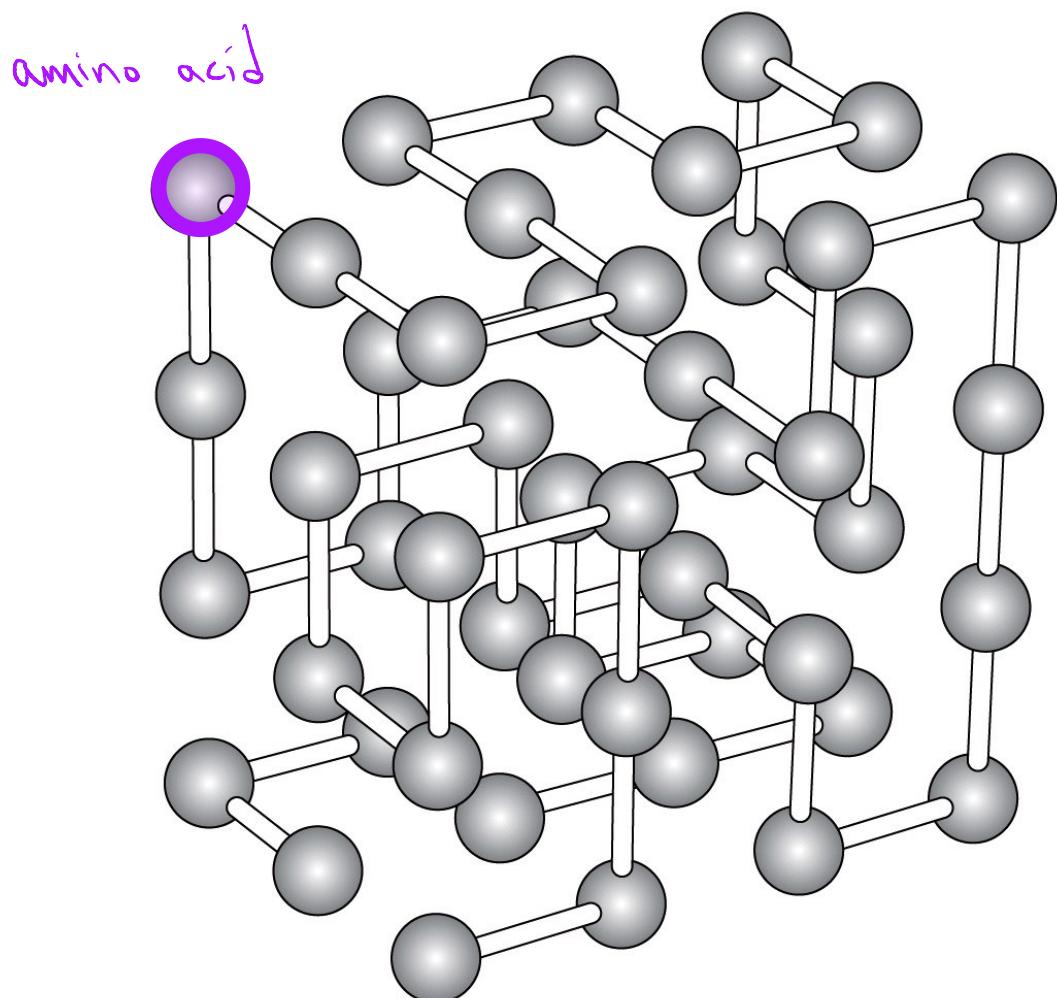
random SAW

# What is a structure?

Proteins as random walks

PBoC 8.4

Simple model: Self-avoiding, compact random walk on a cubic lattice



# What is a structure?

HP model

*consider the solvent, water*

More complex model: Hydrophobic and polar residues

Number of possible 3D conformations is so large that a random search would take a long time:

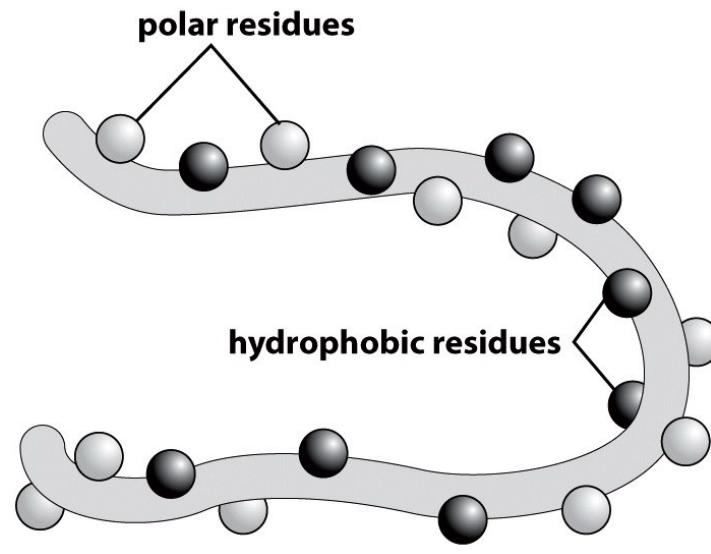
100-monomer chain  
 $6^{100} = 6.5 \times 10^{77}$

One structure per femtosecond  
 $2 \times 10^{55}$  years

Age of universe  $\sim 10^{10}$  years

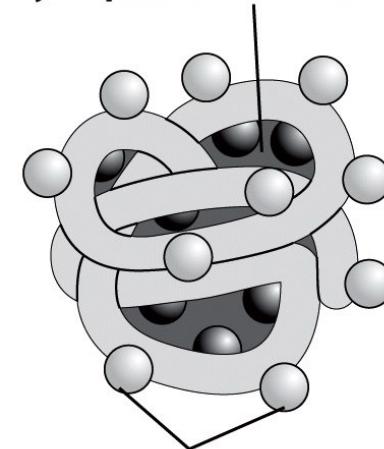
One key: hydrophobicity

20 AA  $\rightarrow$  HP



unfolded polypeptide

free energy lowered by sequestering hydrophobic residues

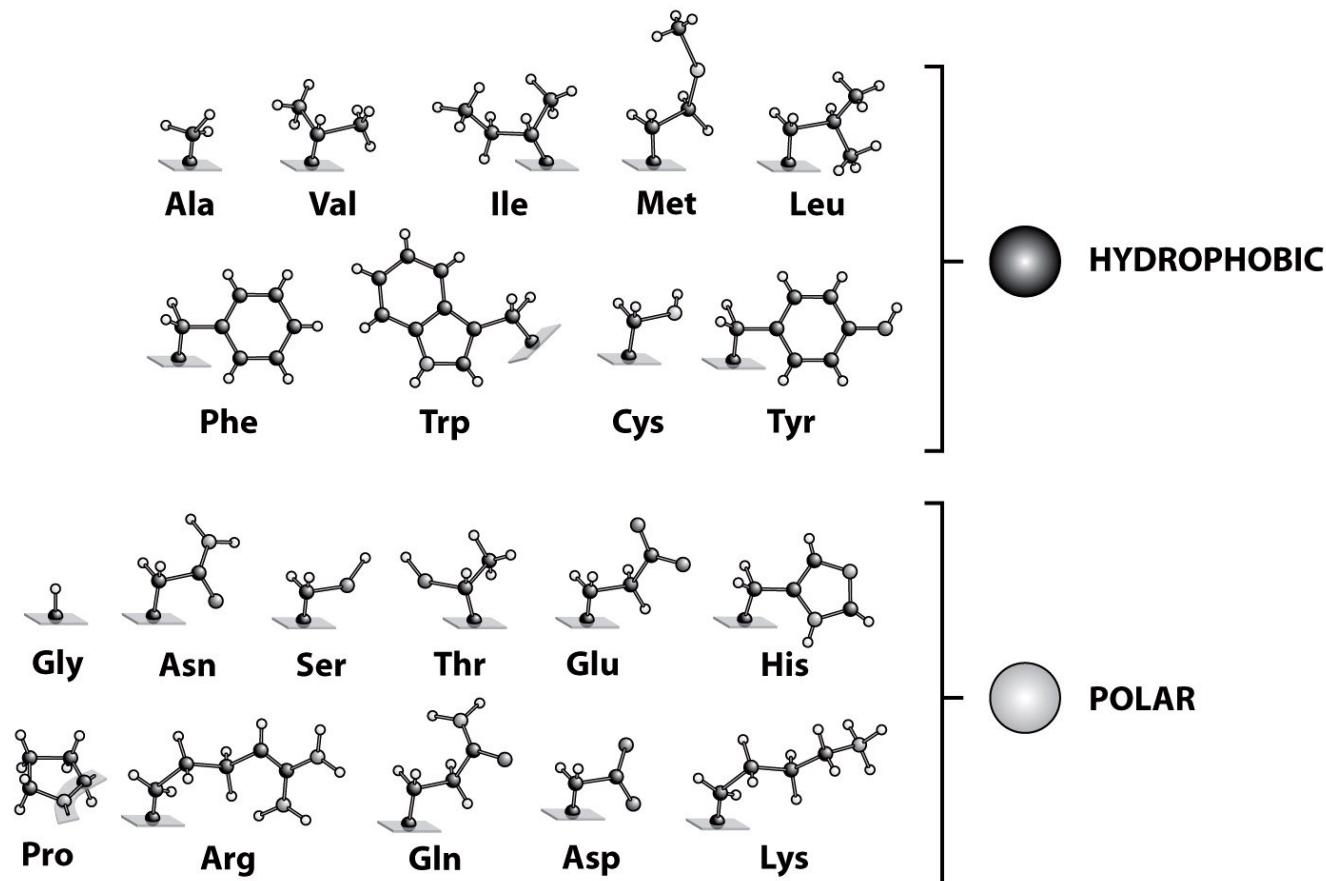


polar residues participate in hydrogen bond network

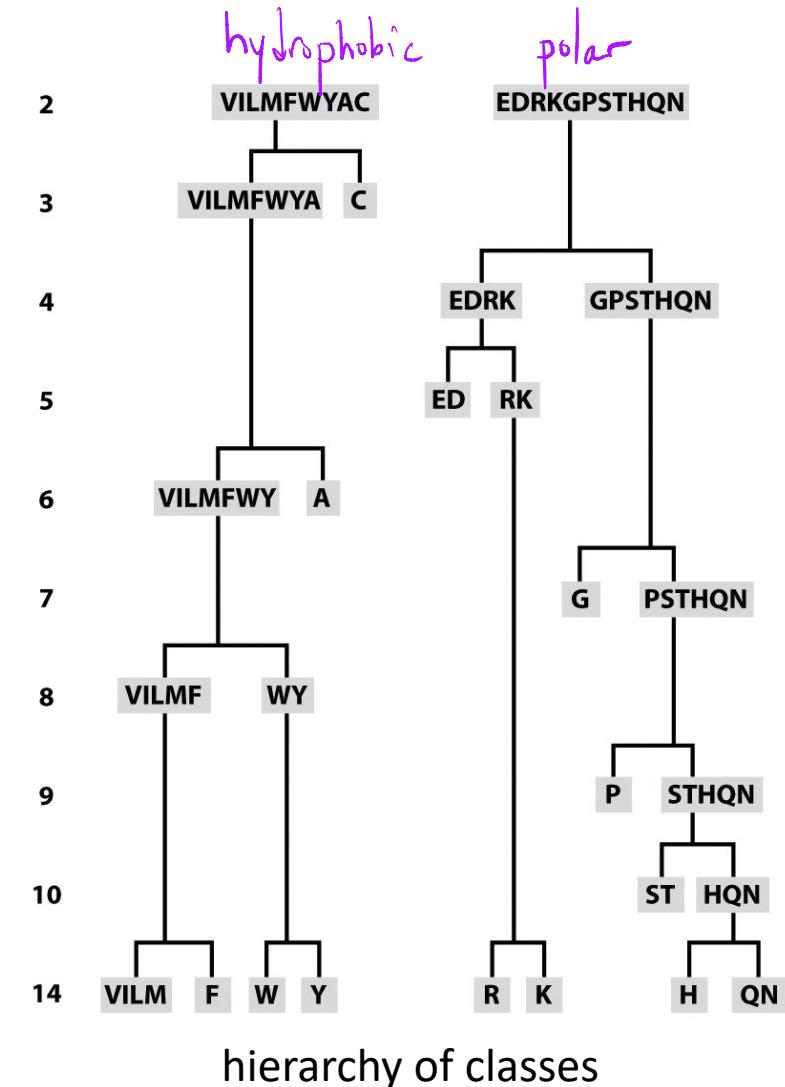
folded conformation in aqueous environment

# What is a structure?

HP model



coarse-grained model: HP



# What is a structure?

## HP model

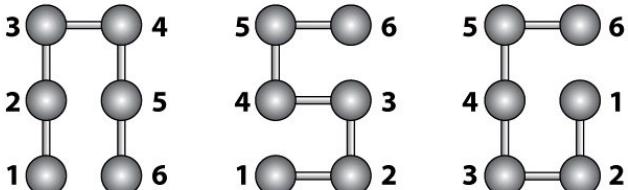
toy HP model:  
6 monomers on a  
3x2 lattice

sequences:  $2^6 = 64$

sequence HPHPHP

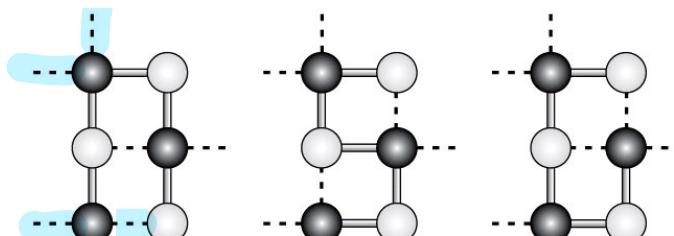
sequence PHPPHP

(A)



number of unique structures: 3

(B)

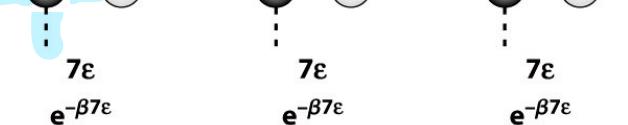


interaction model: assign free energy penalty for H-P or H-solvent interactions (---)

Given an HP sequence, which of the possible structures minimizes the total free energy?

energy  
weights

energy  
weights



$$p_{\text{fold}} = \frac{e^{-2\beta\epsilon}}{e^{-2\beta\epsilon} + 2e^{-4\beta\epsilon}} = \frac{\text{Boltzmann factor for state}}{\text{partition function}}$$

# Protein folding problem

- (i) What is the physical code by which an amino acid sequence dictates a protein's native structure?
- (ii) How can proteins fold so fast?
- (iii) Can we devise a computer algorithm to predict protein structures from their sequences?

:

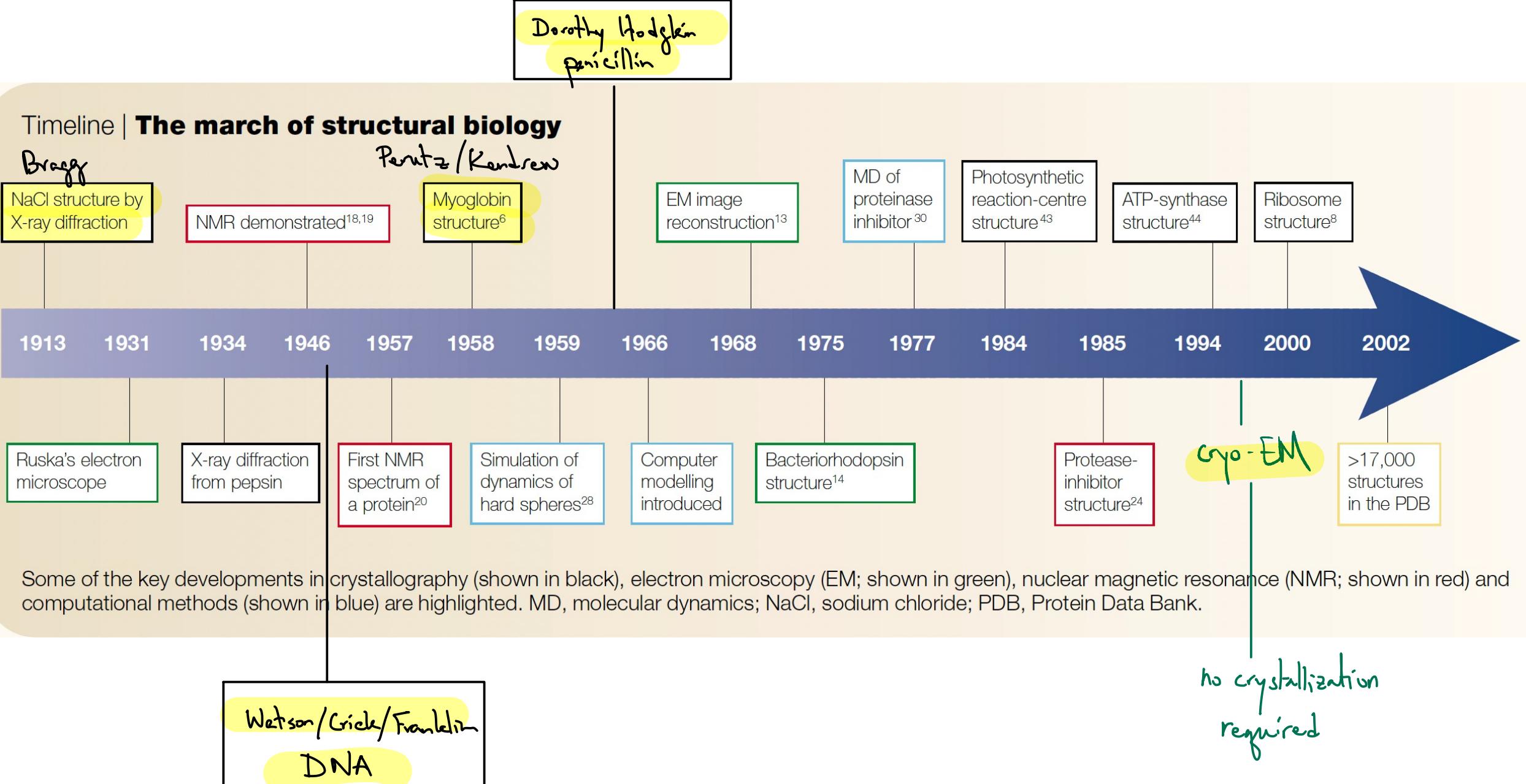
What sequences are protein-like?

# Lectures 3, 4, 5

- Roderick Clayton: Purification of photosynthetic reaction center (1968)
- George Feher: Reaction center structure (1974)  
*"For his many contributions to the understanding of the physics of photosynthesis; specifically, for his role in the pioneering of the concept of reaction centers in photosynthetic bacteria, their isolation, their spectroscopy and their structural characterization."* (1982)
- Peter Wolynes: Protein folding, energy landscape (1987)  
*"For his conceptual breakthroughs in protein dynamics and protein folding, and his critical insights toward the understanding of how proteins work at the most fundamental level."* (2004)
- Jose Onuchic: Protein folding routes (1995)
- Ken Dill:  
*"For independent contributions to a new view of protein folding, from the introduction and exploration of simple models, to detailed confrontations between theory and experiment."*

1895 Roentgen x-rays

2020: AlphaFold



Your mission:

Integrate information about the context of biomolecular structure.

Name one open question in the field of biomolecular structure

a priori structure prediction  $\text{AA} \rightarrow \text{3D structure}$

Understand / explain the key developments in the field

Give an example of a breakthrough described, and identify its "scale"

- X-rays, fueled the field ~ human-scale ~ curiosity / accidental  
→ impact on medical technology
- computational structure determination
- cryo-EM, no crystals needed!

Classify "genre" of information sources

• expository partially Identify the genre used in the media

• argumentative we listened to / watched

• narrative primarily

# Lecture 3: Photosynthetic reaction centers

## Clayton & Feher

- Roderick Clayton: Purification of photosynthetic reaction center (1968)
- George Feher: Reaction center structure (1974)

*"For his many contributions to the understanding of the physics of photosynthesis; specifically, for his role in the pioneering of the concept of reaction centers in photosynthetic bacteria, their isolation, their spectroscopy and their structural characterization."* (1982 Delbrück Prize)