

# 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**

## Why is the course designed this way?

Explore the way science is done and who scientists are

Access resources beyond textbooks, to gain insights and experience

# 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

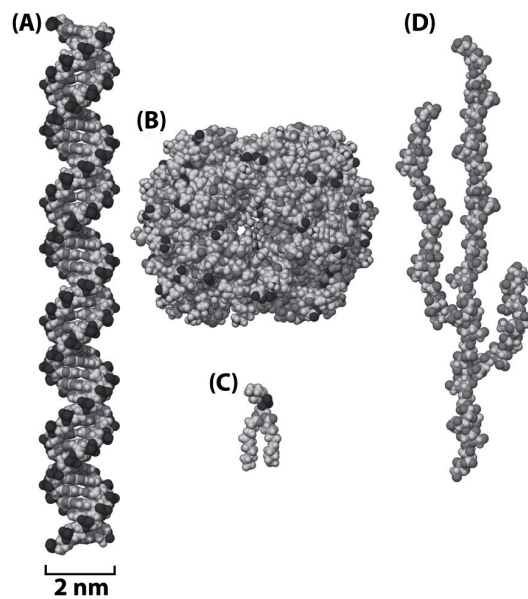
- The stuff of life
- What is a structure
- Proteins as random walks

PBOC Chapter 1.1, 1.2, 8.1, 8.4

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

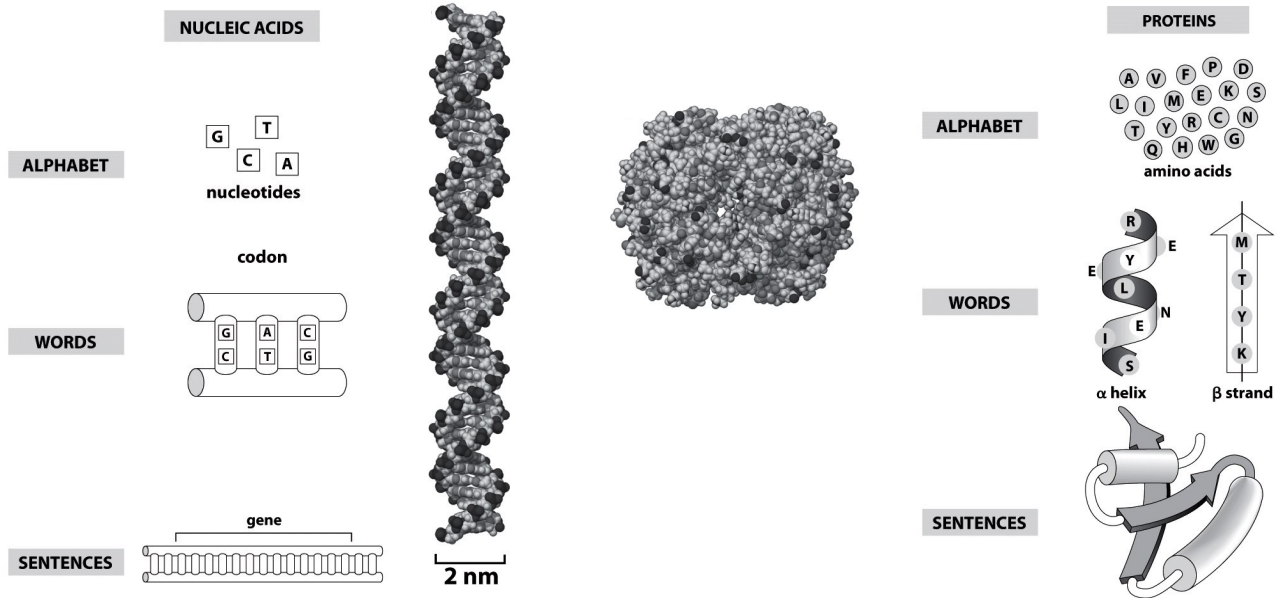
## The stuff of life

What macromolecules are cells made of?



# The stuff of life

Nucleic acids and proteins are polymer languages



## Why structure?

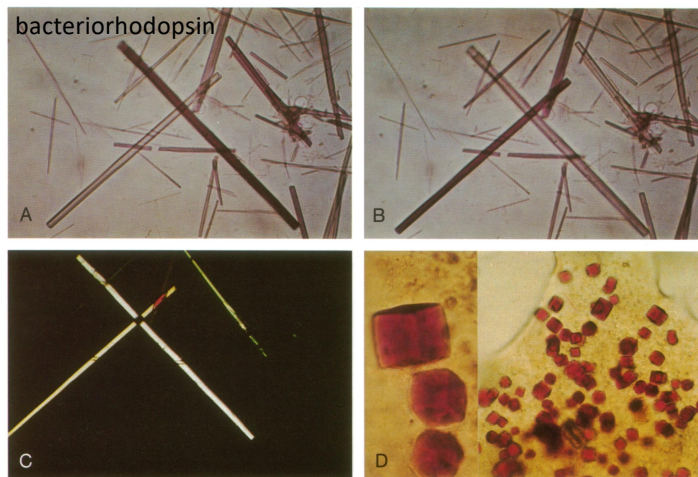
Structure-function relationship

# 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

**RCSB PDB** PROTEIN DATA BANK

201,789 Structures from the PDB  
1,066,577 Computed Structure Models (CSM)

3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM | Help

Advanced Search | Browse Annotations

PDB-101 | PDB | EMDatabank | BIOACID NETWORK | wwwPDB | PDB-Dev

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

General Help

### 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

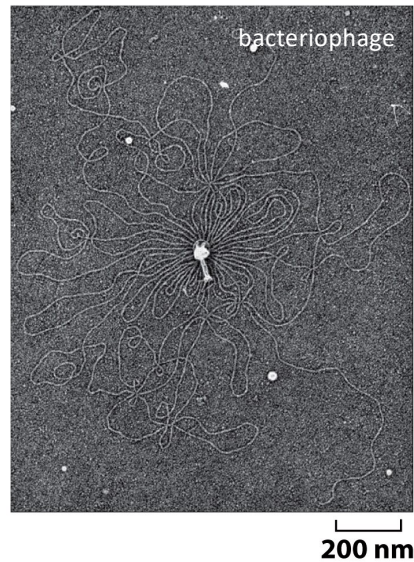
Four levels of hierarchy are commonly used: Entry, Entity, Instance, and Assembly:

- 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

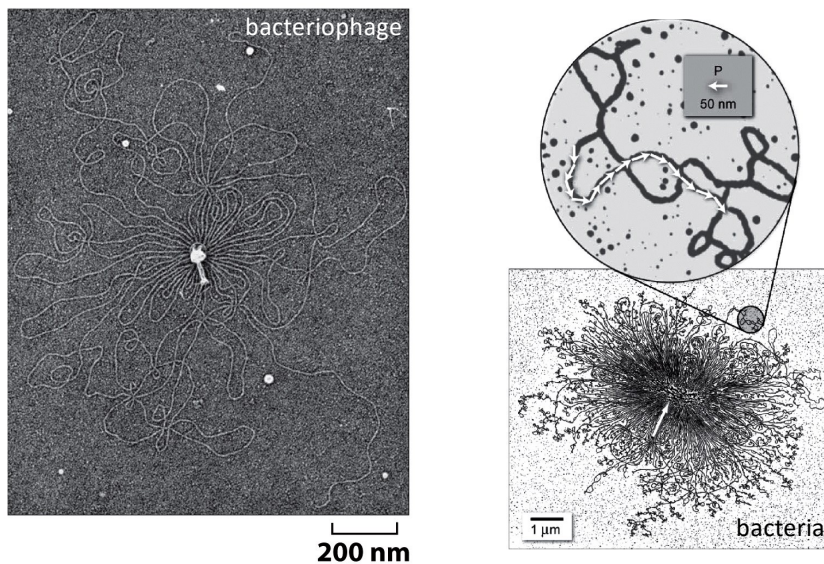
Average size and shape, fluctuations are important



# What is a structure?

Deterministic vs. **statistical** descriptions

Average size and shape, fluctuations are important

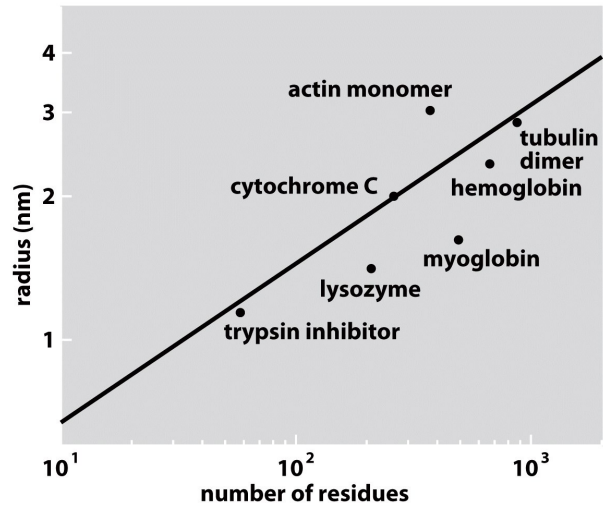
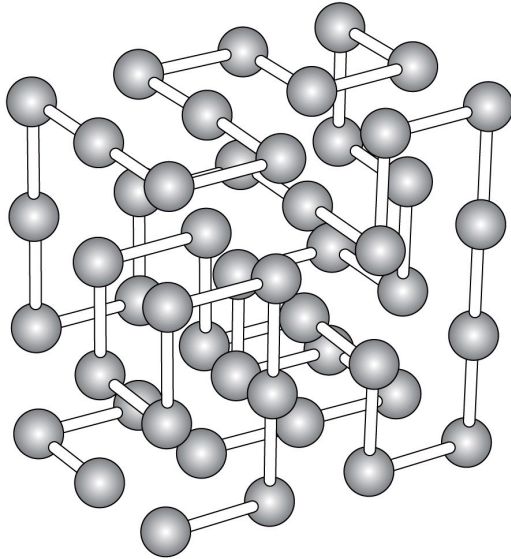




# What is a structure?

## Proteins as random walks

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



# What is a structure?

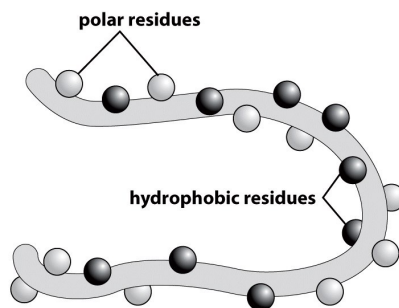
## HP model

More complex model: Hydrophobic and polar residues

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

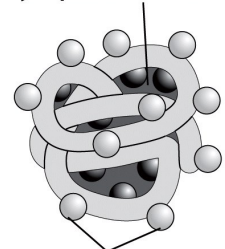
100-monomer chain  
 $6^{100} = 6.5 \times 10^{77}$

One structure per femtosecond  
 $2 \times 10^{55}$  years  
 Age of universe  $\sim 10^{10}$  years



unfolded polypeptide

free energy lowered by sequestering hydrophobic residues

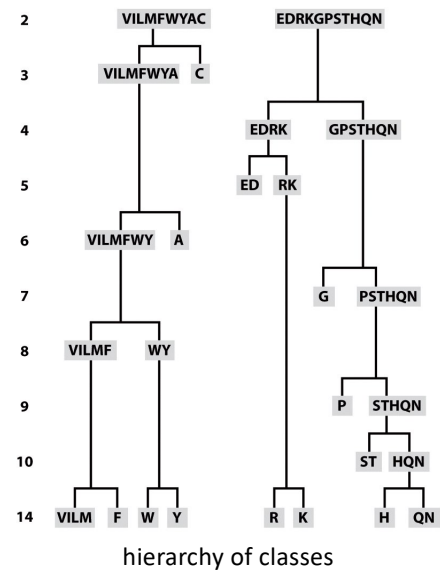
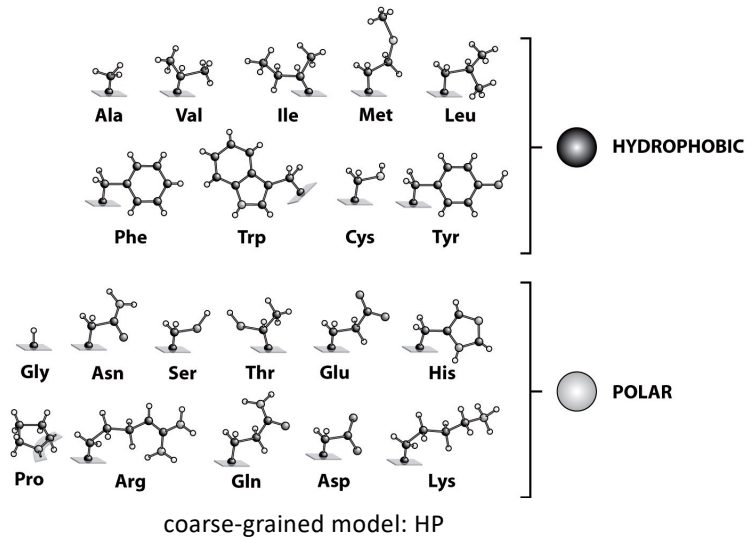


polar residues participate in hydrogen bond network

folded conformation in aqueous environment

# What is a structure?

## HP model



# What is a structure?

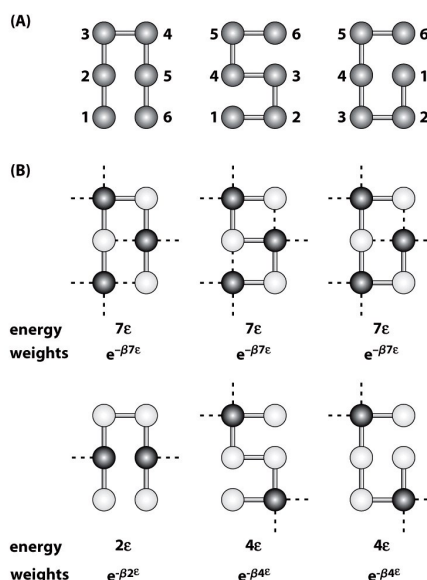
## HP model

toy HP model:  
6 monomers on a  
3x2 lattice

sequences:  $2^6 = 64$

sequence HPHPHP

sequence PHPPHP



number of unique structures: 3

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?

$$p_{\text{fold}} = \frac{e^{-2\beta\epsilon}}{e^{-2\beta\epsilon} + 2e^{-4\beta\epsilon}}$$

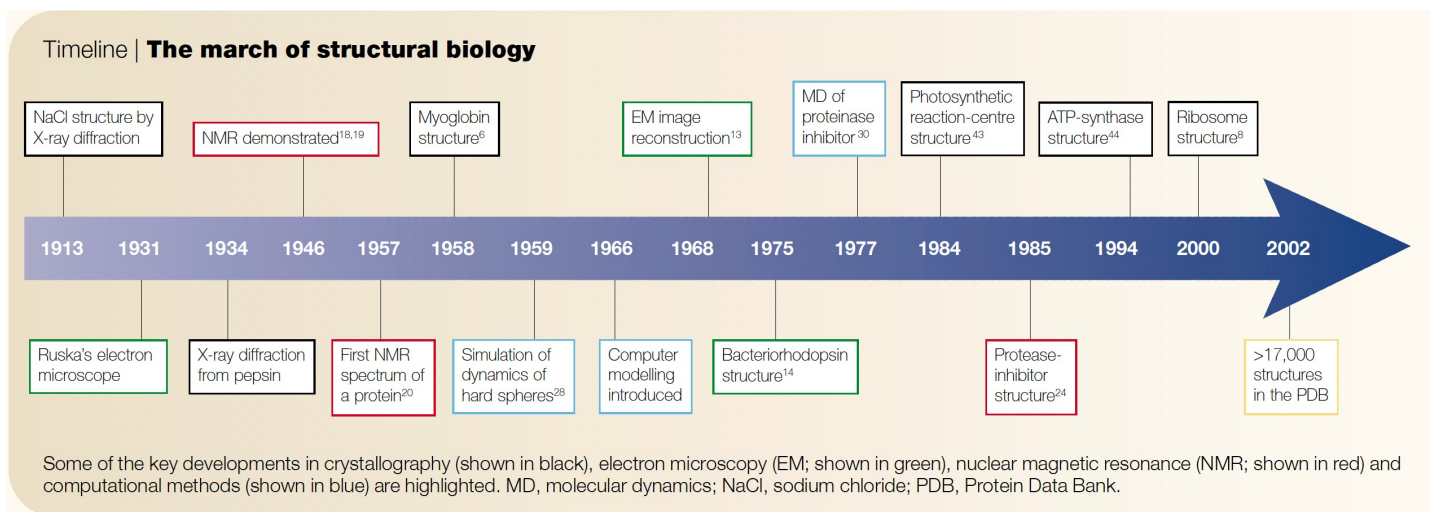
# Protein folding problem

- (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?

## 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."*





# 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 Delbruck Prize)