Computational Biomolecular Design

Motivation and Introduction

Presented by

Dr. Fayyaz-ul-Amir Afsar Minhas

http://faculty.pieas.edu.pk/fayyaz

Department of Computer & Information Sciences
Pakistan Institute of Engineering & Applied Sciences
PO Nilore, Islamabad 45650
Pakistan

Welcome to the Course!

- **Dr. Fayyaz-ul-Amir Afsar Minhas**
  - **Senior Scientist**
    - **Pakistan Institute of Engineering & Applied Sciences (PIEAS), Islamabad, Pakistan.**
  - **PhD Computer Science**
    - **Research: Machine learning in Bioinformatics**
    - **Colorado State University, Fort Collins, Colorado, USA**
    - **Supported by the Fulbright scholarship program**
  - **http://faculty.pieas.edu.pk/fayyaz/**

Quiz 0

- http://goo.gl/forms/Ocn6Qn5Pr6

Logistics

- https://piazza.com/pieas.edu.pk/fall2015/cis533/resources

- **Sign up on Piazza or send me an email to be enrolled online!**

What are the biggest challenges we face?

- We = Humans
- Food
- Environment
- Disease
- War

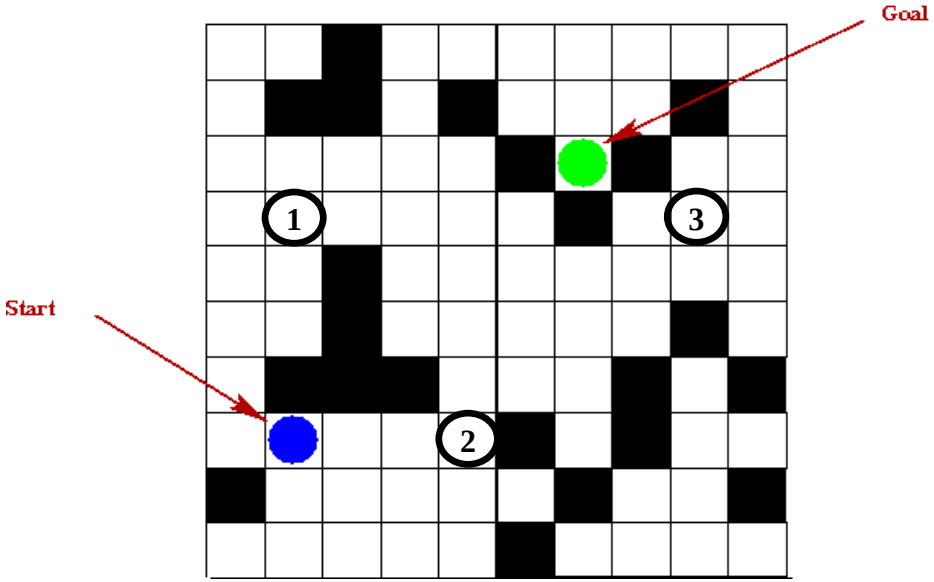
How can we help?

- We = Computer Scientists
- Biology is being transformed by computing
- Computer Aided Design
  - For Biology

Let's start with a puzzle

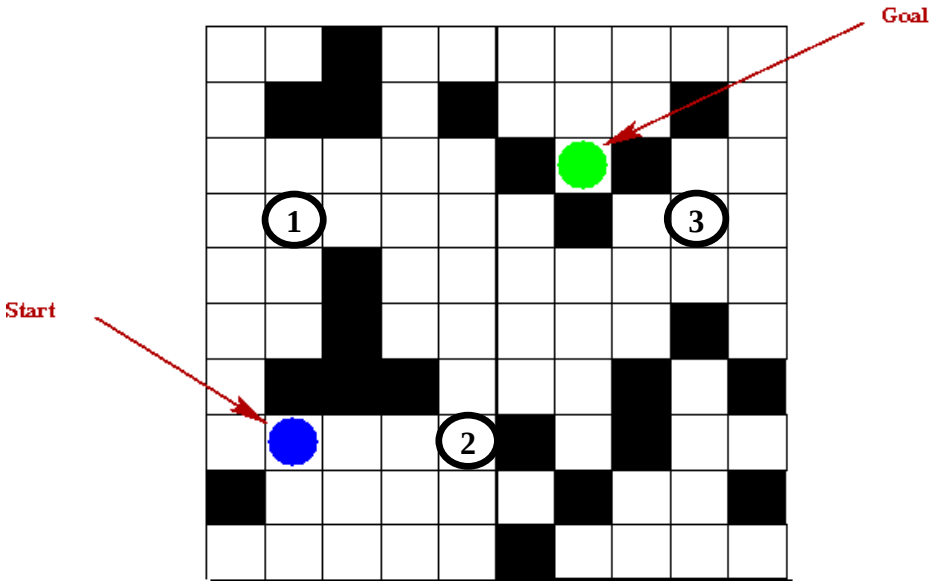
- If you can think about at least one strategy to solve this, your are perfect for this course!
  - We need a strategy for the solution, not the solution!

- Assume you have a robot that can be given an arbitrarily long sequence of 4 commands (N,E,W,S) to make it move on a 2D rectangular grid with obstacles. For example, NNN will cause the robot to move up three steps.
- Your objective is to make your robot reach a given goal point (x_g, y_g), starting from a start point but only after passing through a set of given way points $\{(x_i, y_i) \mid i = 1 \ldots N\}$ in any order.
- However, taking a step consumes energy and you don't want to tire your robot or make it fall off the grid or get eaten by a cougar hiding in the obstacles



Robot Puzzle

- Interesting things about the puzzle
  - If you know a path string, you don't need anything else to plot the path taken by the robot and get its final location
  - However, there can be a large number of possible path strings for a given problem and we need to find the best one



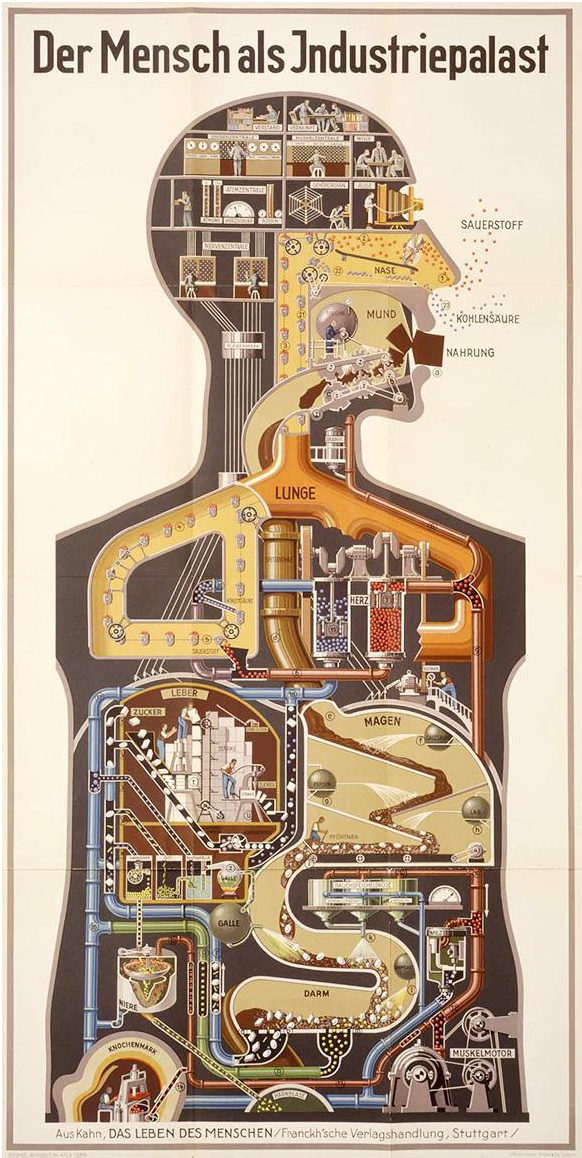
- Remember, solving this puzzle is a sufficient condition but not a necessary one for you to be able to take this course!

- Otherwise you will learn multiple strategies to do this in this course

Be prepared!

- For a very different view on Life
  - We shall view life at the molecular level
  - A different way of thinking – viewing organisms as a chemical factory

- Objective of this lecture
  - Provide motivation for this course
  - Why you should or shouldn't take this course
  - Logistics

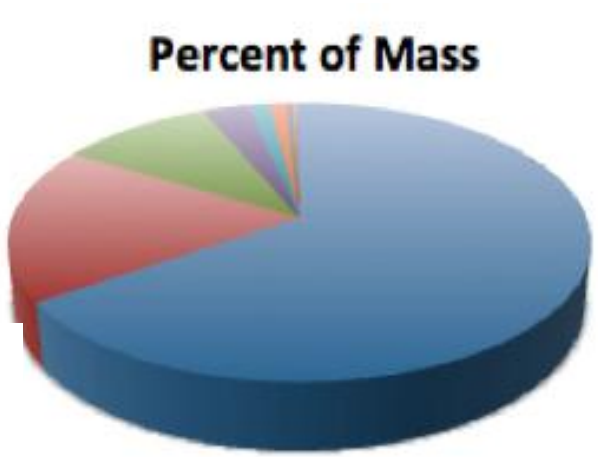


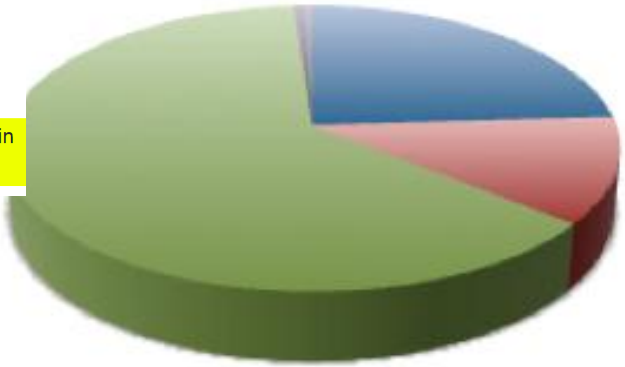
What are we made of?

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	*	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	**	Lr	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Fl	Uup	Lv	Uus	Uuo

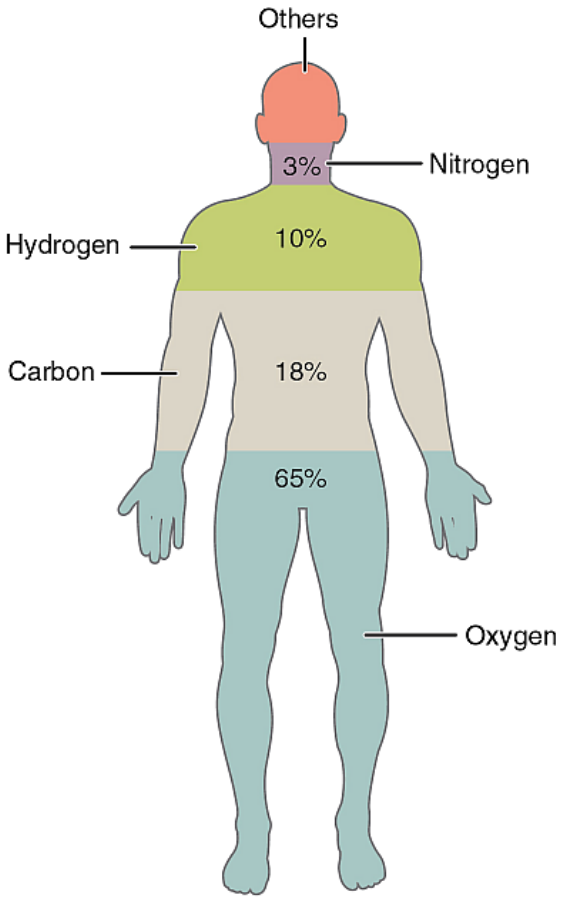
*	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb
**	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No

The four organic basic elements	Quantity elements	Essential trace elements	Possible structural or functional role in mammals





What are we made of?



Element	Symbol	Percentage in Body
Oxygen	O	65.0
Carbon	C	18.5
Hydrogen	H	9.5
Nitrogen	N	3.2
Calcium	Ca	1.5
Phosphorus	P	1.0
Potassium	K	0.4
Sulfur	S	0.3
Sodium	Na	0.2
Chlorine	Cl	0.2
Magnesium	Mg	0.1
Trace elements include boron (B), chromium (Cr), cobalt (Co), copper (Cu), fluorine (F), iodine (I), iron (Fe), manganese (Mn), molybdenum (Mo), selenium (Se), silicon (Si), tin (Sn), vanadium (V), and zinc (Zn).		less than 1.0

https://en.wikipedia.org/wiki/Composition_of_the_human_body

What are Biomolecules?

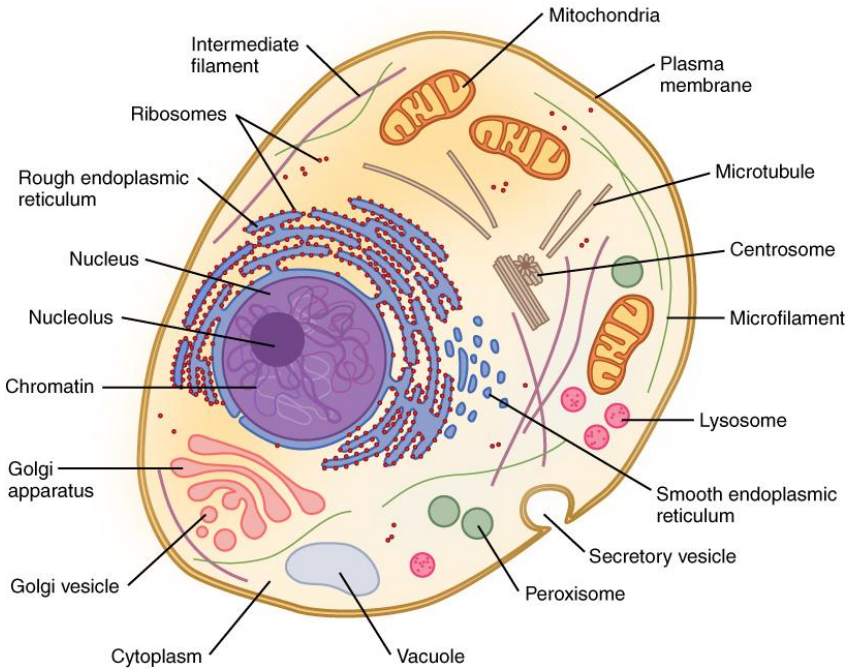
- Any molecule present in a living organism
- Most common?

Molecule	Percent of Mass	Mol.Weight (daltons)	Molecules	Percent of Molecules
Water	65*	18*	1.74e14	98.73
Protein	20	N/A	1.9e10	0.011
Lipids	12	N/A	8.4e11	0.475
Other Inorganics	1.5	N/A	1.31e12	0.74
RNA	1.0	N/A	5e7	3e-5
Other Organics	0.4	N/A	7.7e10	0.044
DNA	0.1	1e11	46*	3e-11

Molecular contents of a typical 20-micrometre human cell

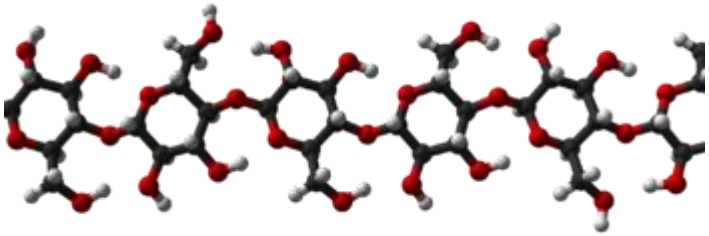
What do these molecules do?

- **Water**
  - **Transport medium (Universal Solvent)**
  - **Regulation of temperature**
  - **Lubrication**
  - **Absorpbtion**
  - **Required: 2-2.5L**
  - **Other molecules cannot function without water**
    - **Proteins are surrounded by and cannot fold without water!**

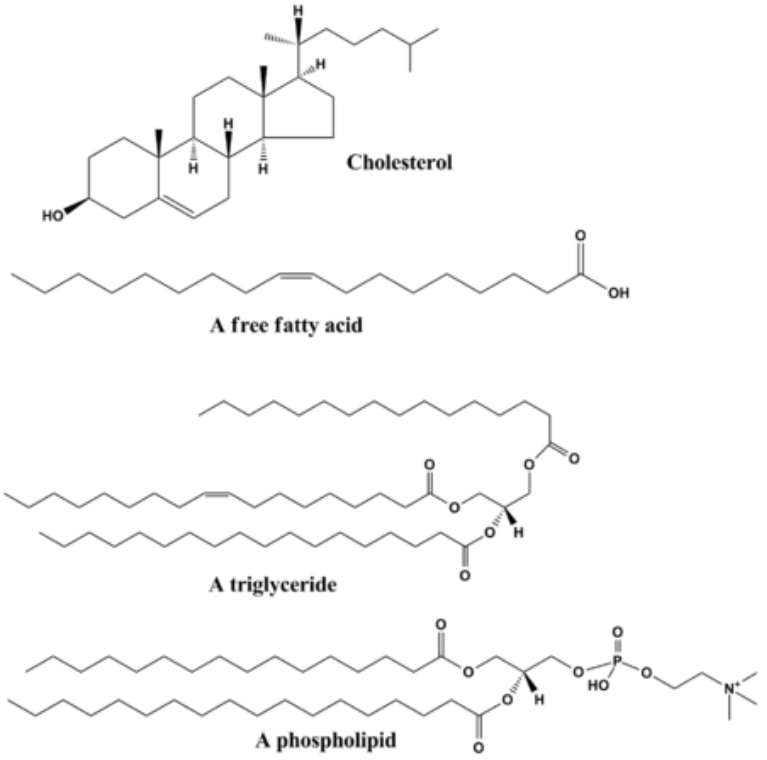


What do these molecules do?

- **Lipids: Store Energy**
- **Polysaccharides: Provide Energy**

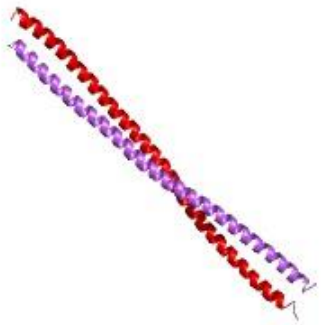


3D structure of cellulose, a beta-glucan polysaccharide.

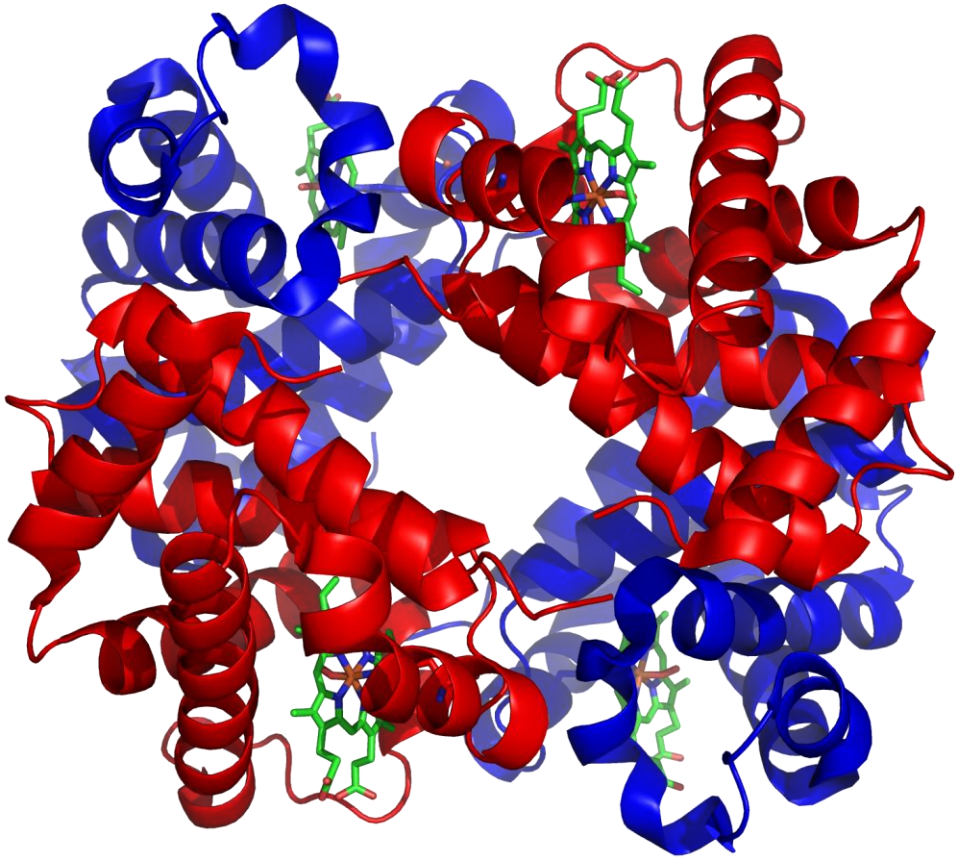


What do these molecules do?

- Protein
  - 50% of the dry weight of the human body is protein
  - Up to 92% of the dry weight of the red blood cell is a single protein called Hemoglobin
  - The key structural material of the outer layer of human skin is a protein called “Keratin”

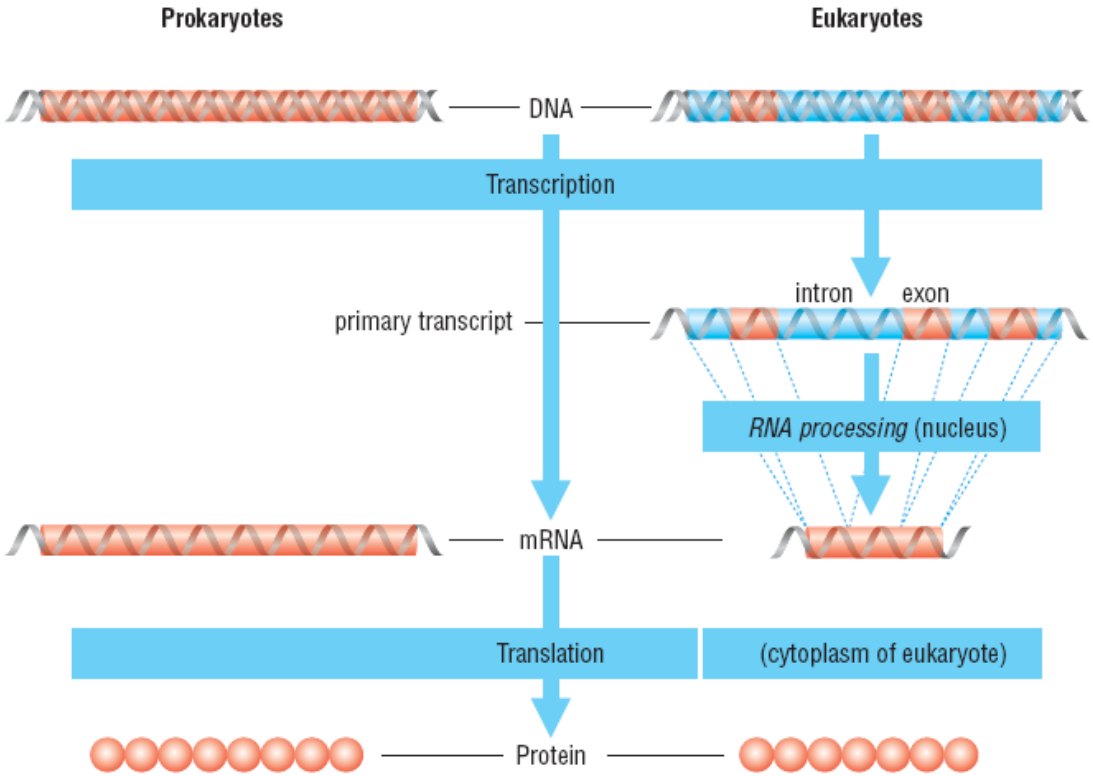


Hemoglobin: A molecule that takes your breath (away!)



DNA and RNA

- DNA: Stores Genetic Information used in creation of proteins
- RNA: A photocopy of the DNA to create proteins (also has other functions0

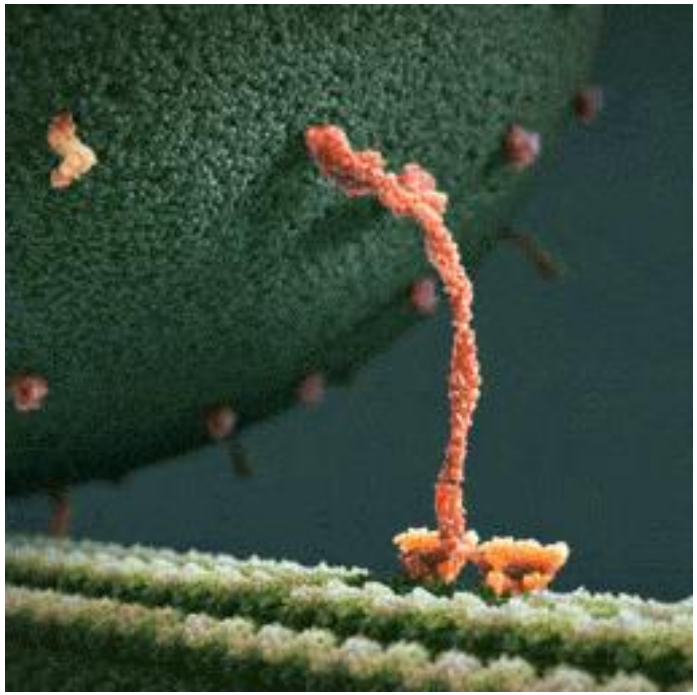


Proteins are the cell's workhorse

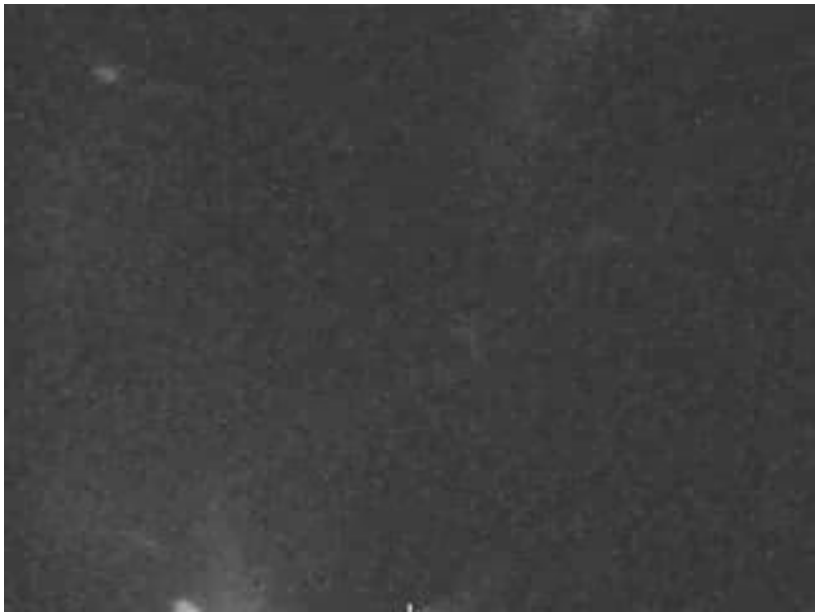
- A protein molecule is a "linear" chain of amino acids
- Each amino acid has some properties which gives proteins very interesting combinations of those properties
- These properties allow the protein to fold and interact with other molecules
- And perform interesting functions!
- All this is very well choreographed!

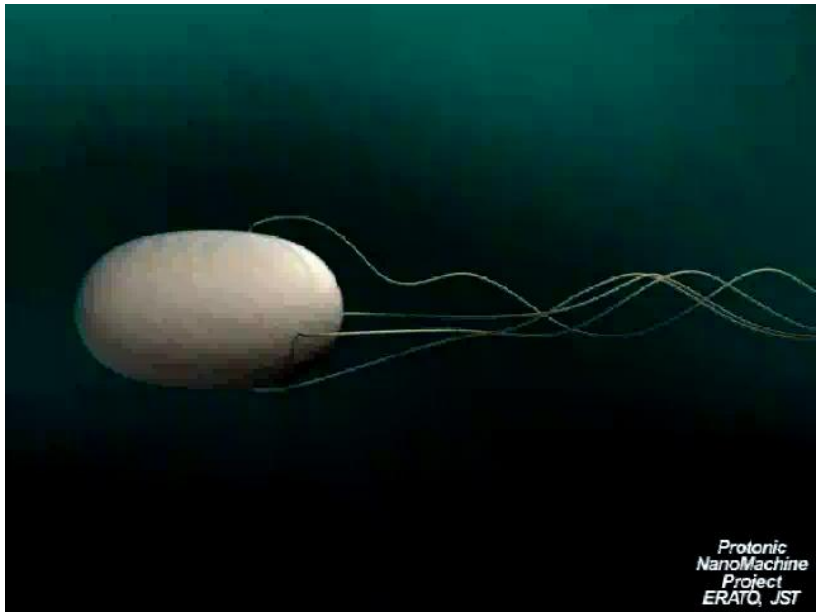
Life is beautiful!

- Kinesin
  - Motor Domain
    - Powered by ATP
  - Tail Domain
    - Attached to “cargo”



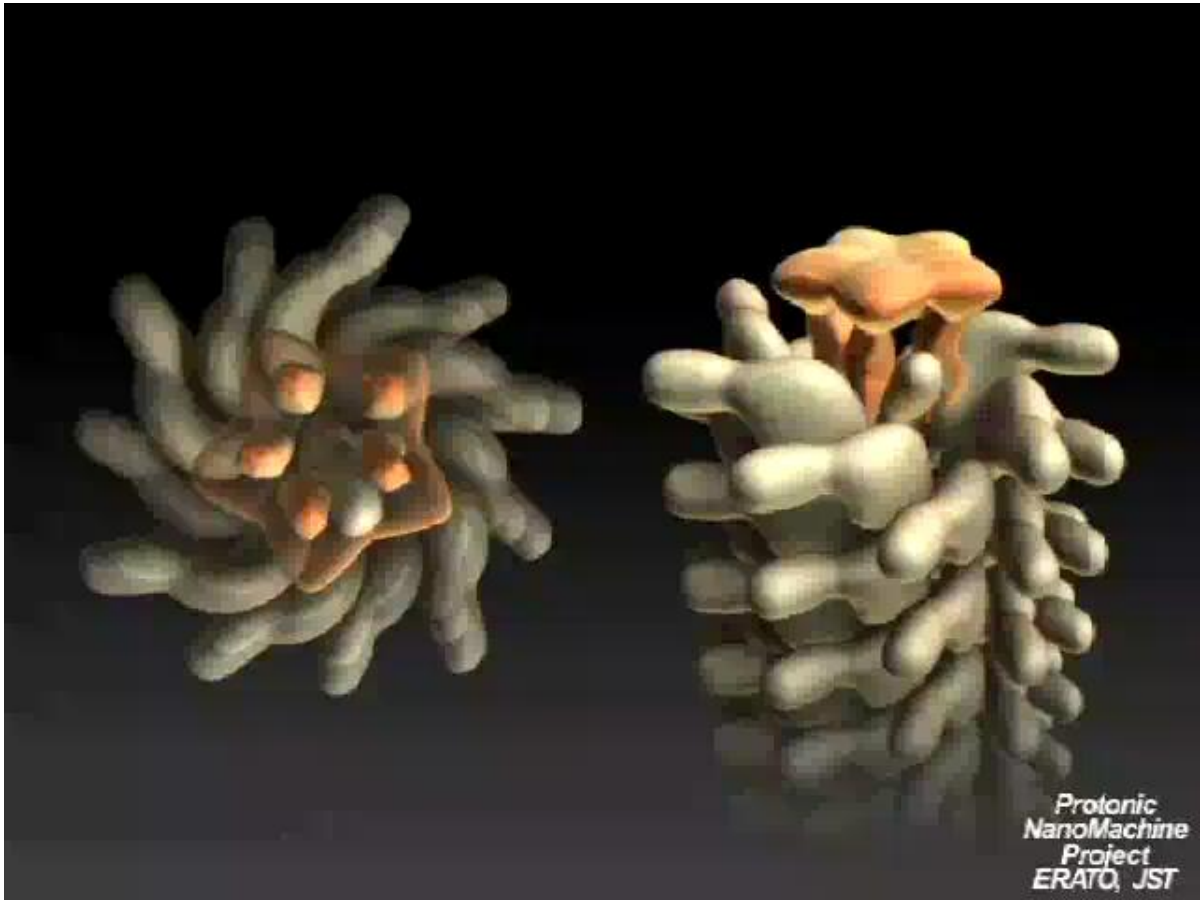
Another machine: Bacterial flagella





http://www.arn.org/mm/mm_movies.htm

Another machine: Bacterial flagella



http://www.arn.org/mm/mm_movies.htm

Proteins are special

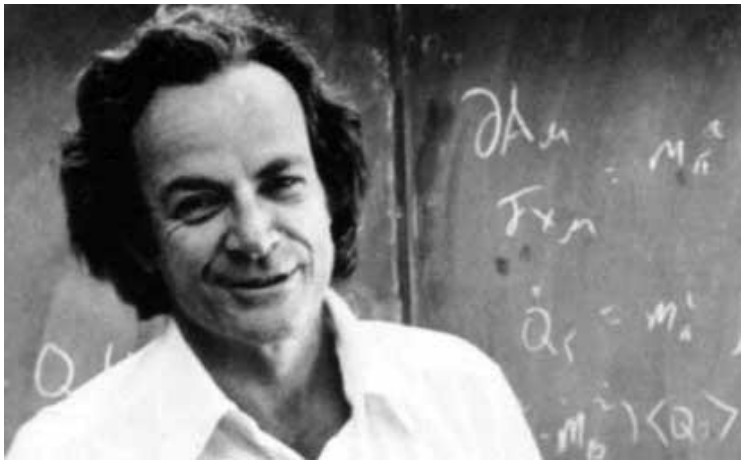
- A protein is a long sequence of amino acids
- A protein folds into a certain (almost always unique) structure based (almost!) solely on its amino acid sequence
  - Anfinsen's Experiment

- So if you can get the sequence of a protein, you can manufacture it

Can we create such machines?

- This idea gave rise to the concept of
  - Synthetic Biology (Life): How to create organisms that do what we want them to do?
  - Biomolecular Design: How can we make biomolecules that do something we want them to do?

- The basic idea came from R. P. Feynman in his talk
  - "There is plenty of room at the bottom"
  - Nanotechnology

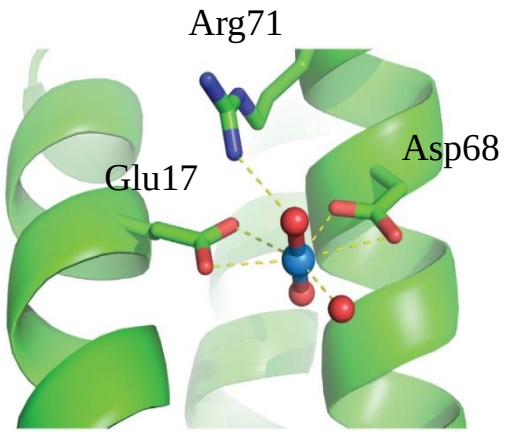


www.youtube.com/watch?v=4eRCygdW--c

What can we do with this?

- The sky (or atom) is the limit
  - Drug design
    - HIV
    - Dengue
  - Nano-robots
  - Oil spill cleanup
  - Increased food supply
  - CO_2 to Diesel
  - Extracting Uranium From Sea Water
- Personally
  - I want to create the world's first Pakistani flag using proteins

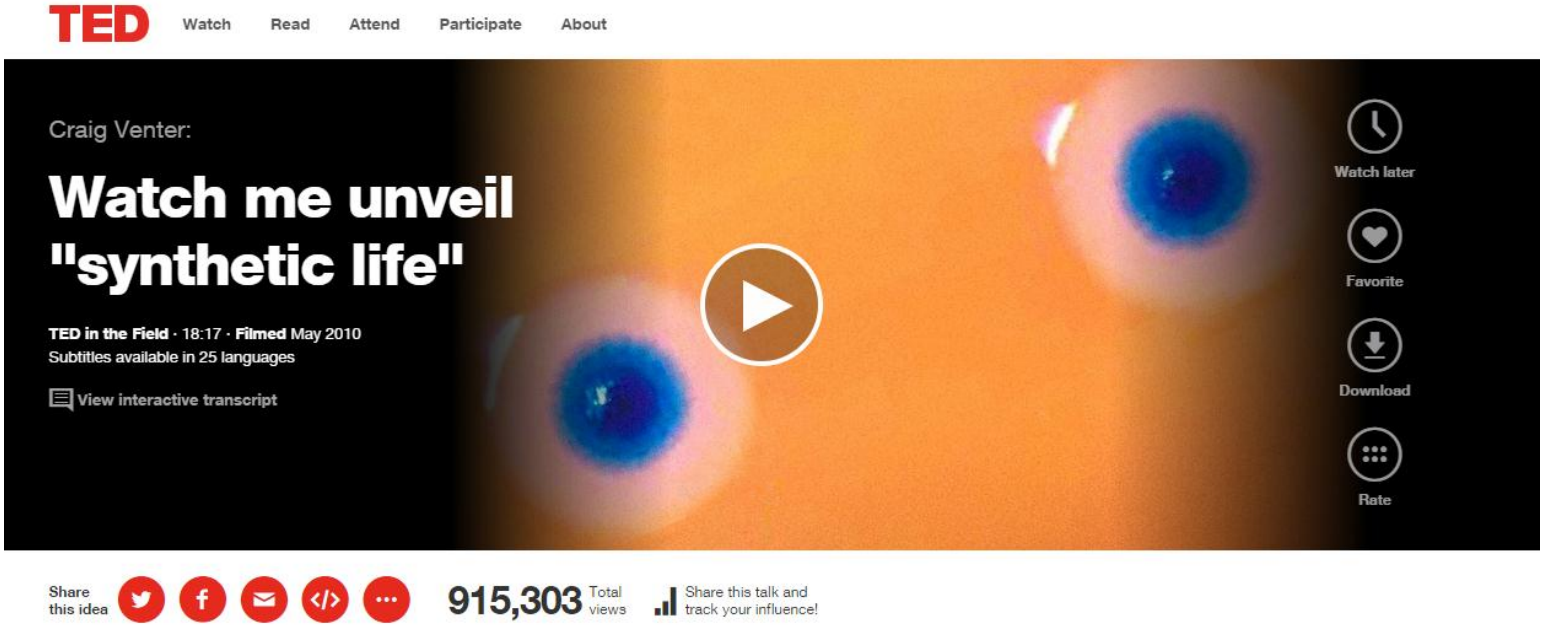
4.5 billion metric tons of uranium, diluted down to a minuscule 3.3 parts per billion



Zhou, Lu, Mike Bosscher, Changsheng Zhang, Salih Özçubukçu, Liang Zhang, Wen Zhang, Charles J. Li, et al. 2014. "A Protein Engineered to Bind Uranyl Selectively and with Femtomolar Affinity." *Nature Chemistry* 6 (3): 236–41.

Where are we right now?

- **Scientists have been able to create a DNA sequence on the compute that has been used to create the world's first synthetic self-replicating bacterial species**



https://www.ted.com/talks/craig_venter_unveils_synthetic_life

How is this a computational problem?

- To understand it thoroughly we need a little more understanding of biochemistry and its underlying physics
  - Structural Proteomics
  - Molecular Dynamics
- Will start in next lecture
- Brief Story
  - For proteins
    - Sequence Drives Structure Drives Function

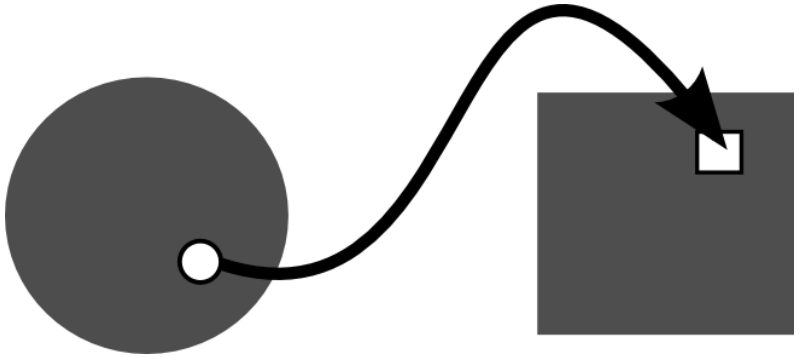
How is this a computational problem?

- To have a certain function the protein needs to have a certain shape
- So the objective of protein design is to make a protein have a certain shape
- However, this requires finding a sequence that folds into a certain shape

- So all this is very much analogous to the robots example
- Both are discrete search and combinatorial optimization problems

Sequence Space Structure Space

Protein Folding



Inverse Protein Folding / Protein Design

What is this course about?

- Artificial Intelligence Meets Biology
- Topics
  - Primer on molecular biology and proteins
  - How do we visualize proteins?
  - How do we model the structure of proteins?
    - Thermodynamics of folding Force fields and Scoring
  - How can we predict the structure of proteins?
  - How do we optimize the structure of proteins?
    - Combinatorial optimization algorithms
    - Introduction to pyRosetta and SHARPEN
  - How do we design new proteins?
    - Using pyRosetta and SHARPEN
  - How do we model dynamics and interfaces of proteins?
    - Docking, Using GROMACS
  - Protein Design Concepts

What is this course about?

- Tools / Technologies used in this course
  - Python
  - PyMOL and VMD
  - pyRosetta
  - Simple docking
  - Molecular Dynamics Simulations using GROMACS
  - Running codes on GP GPUs

Why should you take this course?

- It's fun!
- Very interesting research opportunities
- Interdisciplinary
- Applied Computing where its needed

Why should you be interested?

- Oh, but it involves Chemistry
- Oh, but it involves Biology
- Oh, well there isn't a job market for this
- Oh, well what should I do with this course in Pakistan?
- Oh, well it seems hard
- Oh, well …

Beginning Python

- Resources
- Tutorials

What I cannot create, I do not understand!

(R. P. F)