Structural Bioinformatics & Computational Biology

Computational Biology – 2020

What's it about ?

Computational analysis and study of the structure and function of biological molecules (macromolecules and small molecules)

- Computational analysis and study of biological molecules (macromolecules and small molecules)
- Storage and management of structural information
- Computational algorithms for structure search and comparison
- Computational algorithms for the prediction of 2D and 3D structures of macromolecules
- Computational algorithms for the ***simulation*** of the physical and chemical behavior of biological (Biomolecular Simulation)
- Computational algorithms for protein-ligand analysis and discovery of new drugs (Computational Drug Design)

What will happen?

Learning

- Introduction of structural databases and formats of molecular representation
- Introduction to methods of macromolecule structure determination
- Prediction of secondary and tertiary structure of proteins
- Biomolecular Simulation Methods
- Introduction to protein-ligand interactions, computational docking and virtual screening

Doing

- A little Python and how to use it for programmatic access to structural on-line services and analysis of molecular structures
- How to use the molecular visualizer PyMOL to analyze, display and build structures of molecules (and make pretty pictures!)
- Constructing a protein model by comparative (homology) modelling
- Running a docking simulation with the Autodock Vina software

Who the hell am I ?

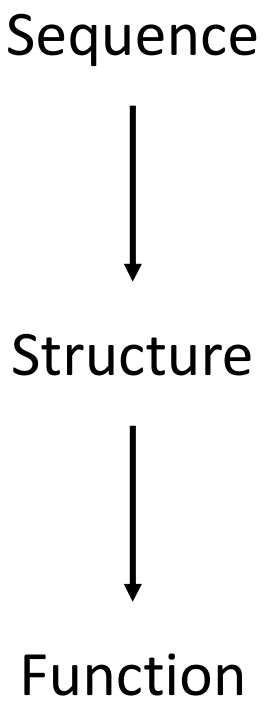
- My name is Paulo Martel
- I teach Bioinformatics, Computational Biology, Drug Design, Protein Structure and Enzymology (far too much!)
- My research focus around computational analysis of protein structure and function using biomolecular simulation techniques, namely Molecular Dynamics, Comparative Modelling and Protein Electrostatics
- My office number is 3.12 in Building 8 (but you probably won't go there…)
- My email is pmartel@ualg.pt
- Feel free to contact me or ~~drop by~~ if you have any questions regarding the course and its content
- This year it will be all remote learning, but I will be ready to answer all your questions using remote means of communication.

Before we start

- Most of computational activities will take place on a **remote server** with address http://compbio2020.ddns.net
- You can login to that server from any computer you wish
- There are accounts for you on this server using, with login name same as your student number (without the leadin “a”).
- Login in with your chosen account name and password **#123bc<your_number_in_reverse> …** we can change those passwords later if you prefer.
- The environment we are using is called Jupyter Hub and provides access both to a Python Notebook interface and one or more login terminals on the server
- You will be required to install the following softwars on you personal computer:
  - PyMOL (you can download it from the site https://www.pymol.org)
  - MGLToosl (the instructor will provide a link for downlad)

Biomolecular Structures

Structure Leads to Function



Flow of Biological Information

Gene …TTAATAAGT…

⬇ transcription

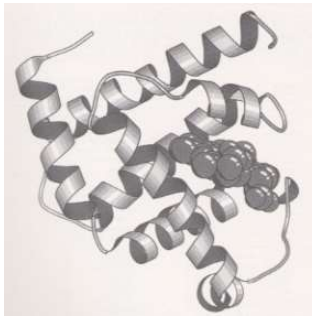
m-RNA …UUAAUAAGU…

⬇ splicing, translation

cadeia polipeptídica …LISVHDN…

⬇ post-translational modifications

proteína



Central Dogma of Molecular Biology

Exceptions: RNA viruses, prions, ribozymes (?)

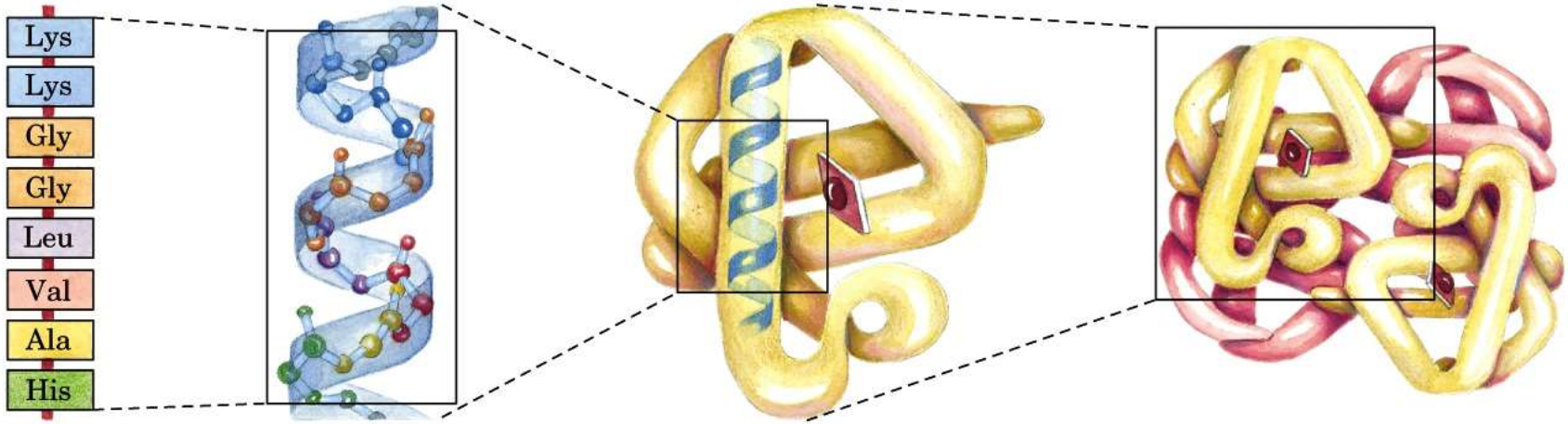
Levels of structural organization

Primary structure

Secondary structure

Tertiary structure

Quaternary structure



Aminoacid sequence

α-helix

Polypeptide chain

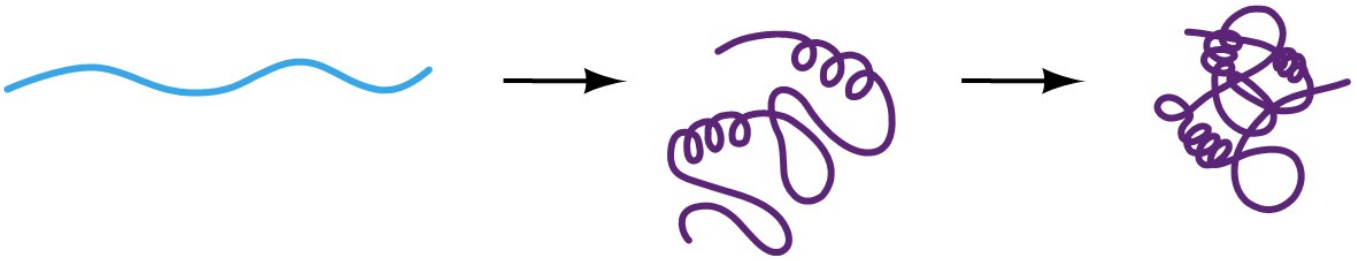
Arrangment of monomers

Sequence determines structure

The tridimensional structure of proteins arises due to the physico-chemical forces acting between atoms in the polypeptide chain and the solvente. Many proteins will spontaneously acquire their *native* structure following ribossomal synthesis. This process is called *protein folding*.

Predicting the native structure of a protein given its sequence is one of the most fundamental probblems of nodern day molecular biology.
(Folding problem)

Protein folding:

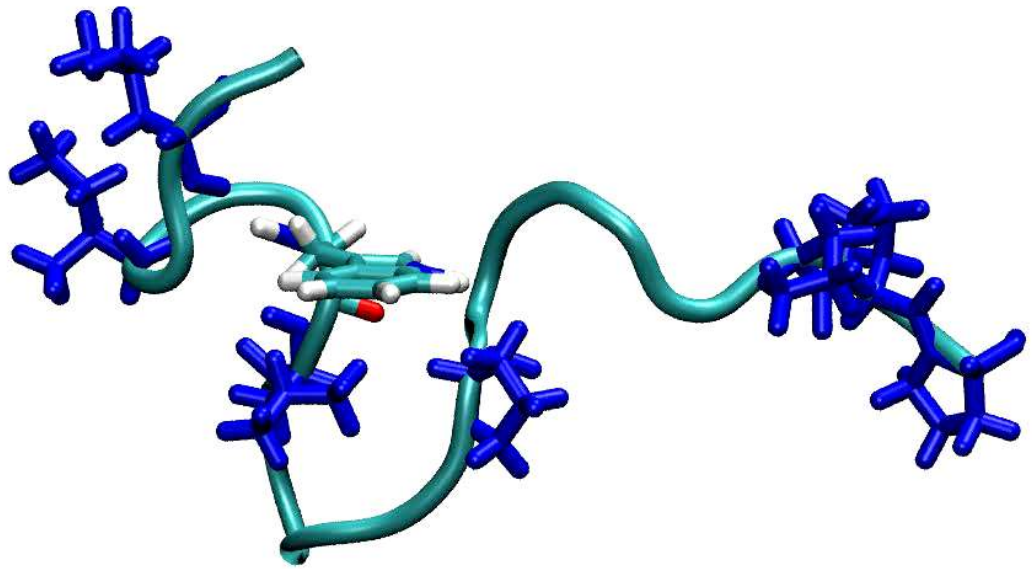


Linear protein chain

Secondary structure formation

Bending of the protein chain in a 3D shape

Computing folding



Molecular Dynamics simulation of the folding mechanism of the mini-protein ***trp*-cage**

Representing Structures

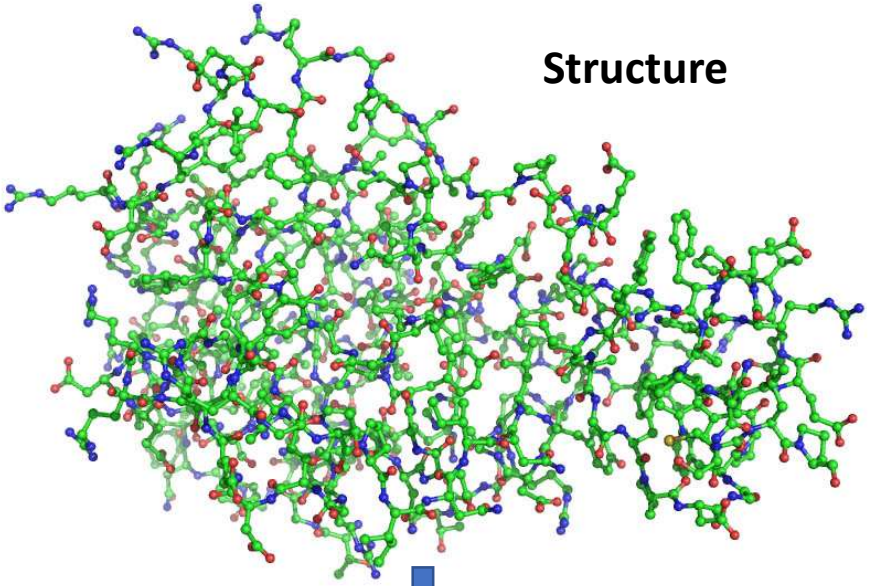
The representation of molecular structure is much more complex and storage intensive than sequence

Sequence

```
...AVAGGATILVHNQDAGEPAIVLAFG...
```

Simple sequence of one-letter symbols

Structure



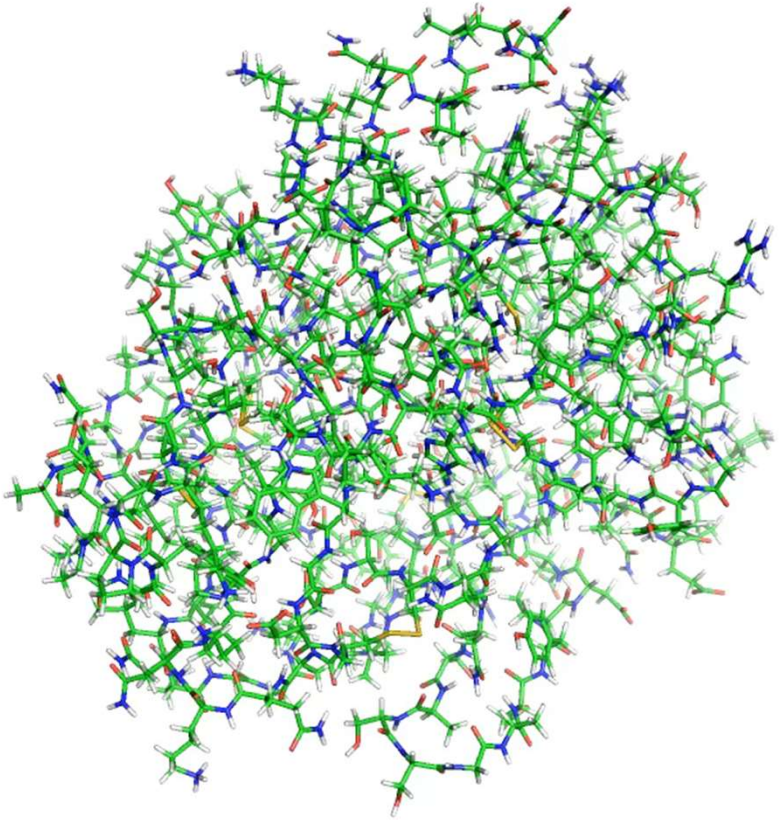
XYZ **coordinates** of each atom, their **types** and **connectivity**

Example

Human Trypsin

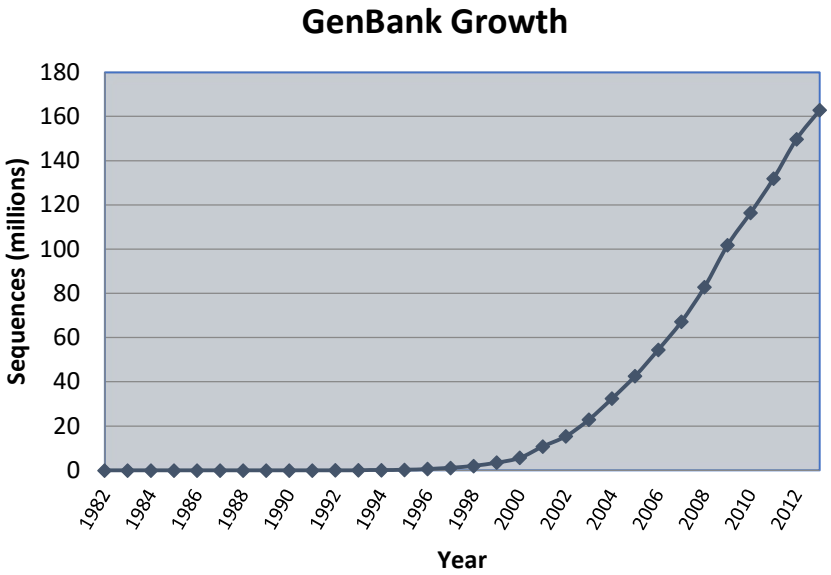
```
>sp|P07477|TRY1_HUMAN Trypsin-1 OS=Homo sapiens OX=9606 GN=PRSS1
MNPLLILTFVAAALAAPFDDDDKIVGGYNCEENSVPYQVSLNSGYHFCGGSLINEQWVVS
AGHCYKSRIQVRLGEHNIEVLEGNEQFINAAKIIRHPQYDRKTLNNDIMLIKLSSRAVIN
ARVSTISLPTAPPATGTKCLISGWGNTASSGADYPDELQCLDAPVLSQAKCEASYPGKIT
SNMFCVGFLEGGKDSCQGDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIK
NTIAANS
```

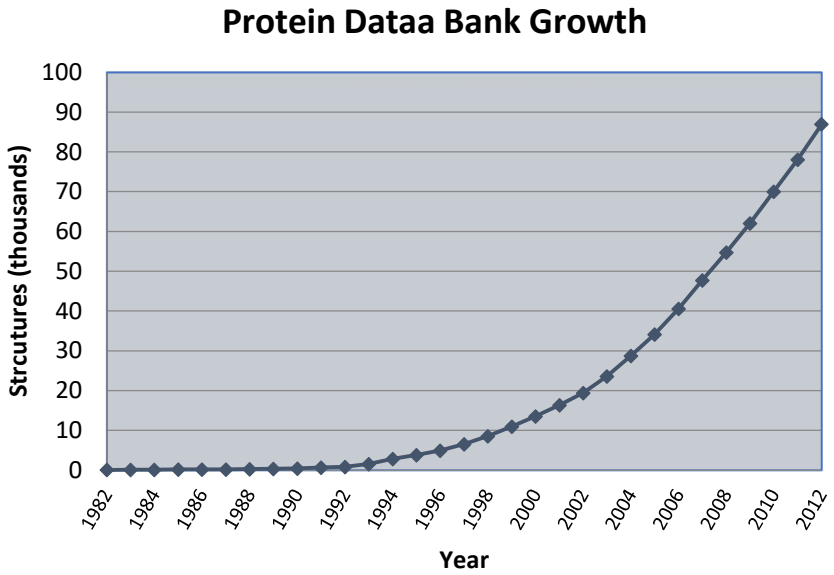
247 aminoacids



3415 atoms

Sequence *versus* structure





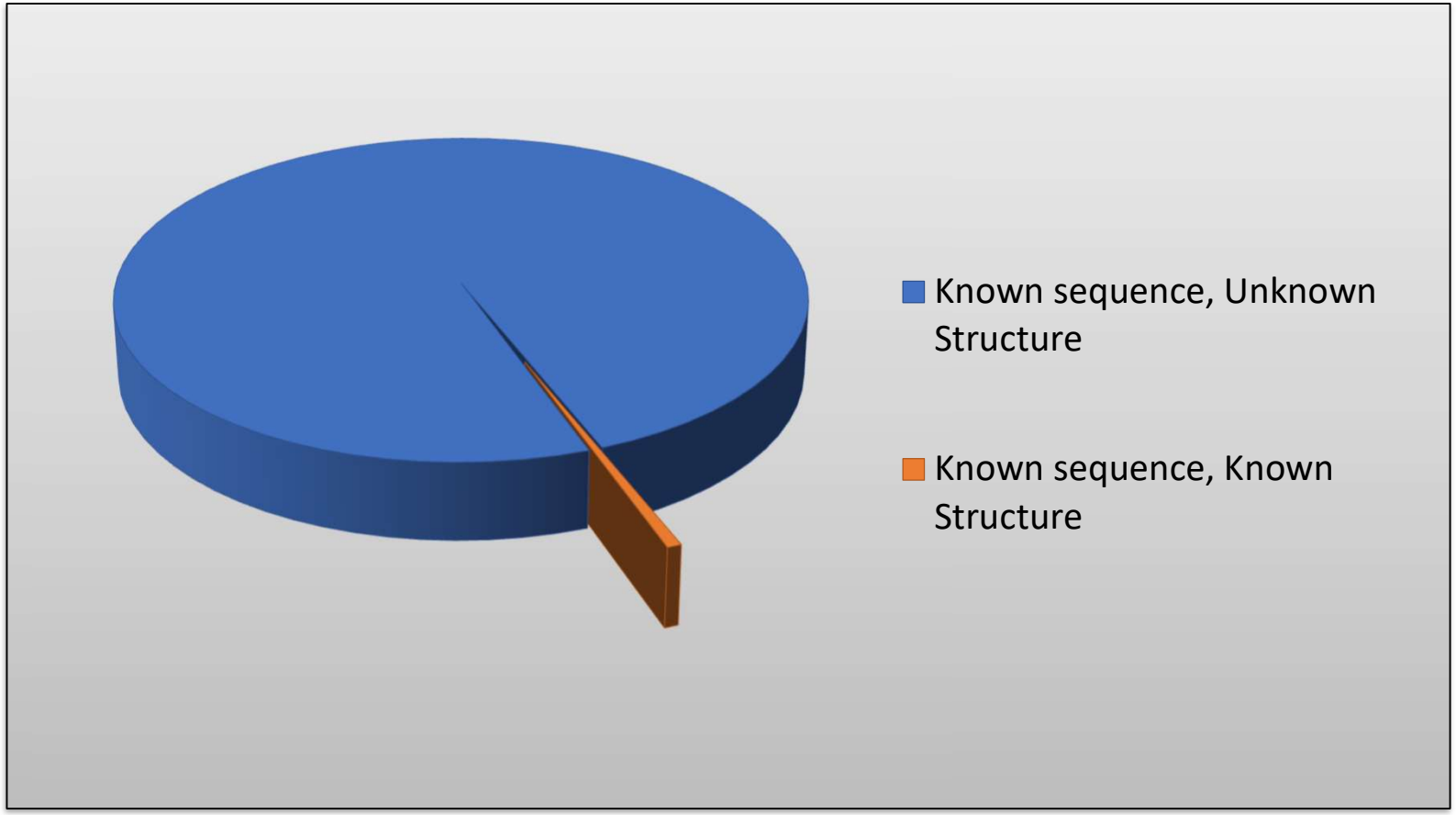
milions of sequences *versus* **thousands** of structures!

In 1982: 172 know structures and 602 sequences

Today (March 2019): 149,886 structures and 212,260,377 sequences!!

In conclusion: Sequencing is way faster than structure determination (the number of proteins of known sequence and unknown structure is growing very rapidly)!

Most knowproteins have unknow sequence !



The importance of structure prediction

The vast (and steadily growing) number of proteins of unknown structure puts a heavy demand on ever faster methods for 3D structure determination. Due to their intrincacies, such methos simply cannot cope with the fast pace of sequencing, and the gap widens more and more. This situation is not likely to change, ever.

So what can we do ?

We need to be able *predict* the 3D structure of proteins from their aminoacid sequence. In general terms this is a very hard computational problem, but there are special situations where it is very feasible. That is currently our best hope of coping up.

Prediction of the 3D structure of proteins is thus one of the most fundamental problems of bioinformatics / computational biology.