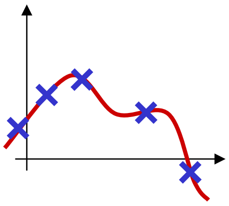
Software Tools for Computational Biology

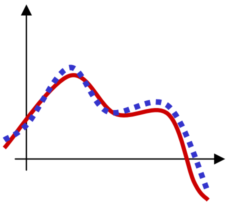
Course introduction

Alessio Bechini

Biology & Computer Science



$y = f(x)$

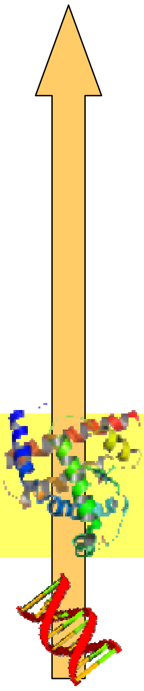


- Biologists focus on experiments, specifics and classifications
- In ordinary Biology, research goes through collecting (x_i,y_i) pairs – i.e., facts – and classifying them. What about f?

- Computer scientists focus on abstract models, algorithms, and tools to work on them
- In Computer Science, interest is towards f (i.e. the overall model in accordance with evidence) and effective/efficient ways to handle-compute it; single pairs (x_i,y_i) are deemed irrelevant

Computational Biology is aimed at investigating biological phenomena by means of tools from Computer Science.

Bio&CS at Different Scales



• Body	• SW and robotics for Surgery
• Tissues/Organs	• Simulations for BioMechanics
• Cells	• Simulation of cell interaction
• Molecules	• Molecular structures, motion and interactions
• Genes	• Sequence analysis

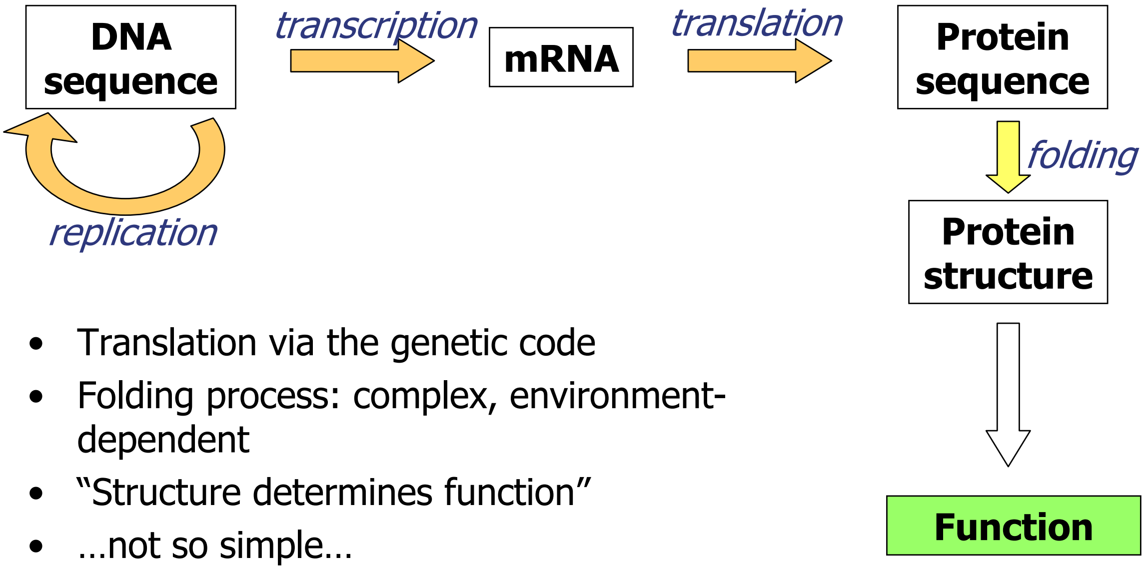
Class Topics

- Scripting in Python
- Molecular modeling
  - Molecular Mechanics
  - Molecular Dynamics
- Structural Computational Biology
  - Proteins
- Docking
- Notes on System Biology

Class Goals

- Overlook of main themes in Computational Biology
- Master current assorted SW tools
  for working on biological problems
- Integrate different tools
  to work on a specific problem
- Be able to collaborate
  in the design and implementation of new tools
  for Computational Biology

Central Dogma of Molecular Biology



DNA sequence

transcription

mRNA

translation

Protein sequence

replication

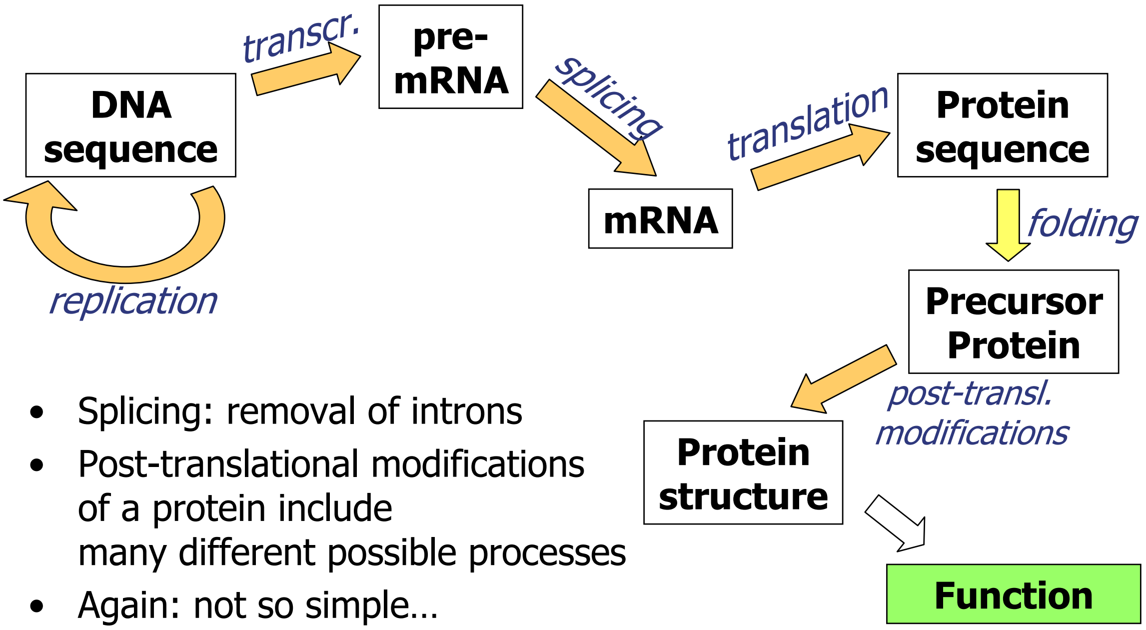
folding

Protein structure

Function

- Translation via the genetic code
- Folding process: complex, environment-dependent
- “Structure determines function”
- …not so simple…

Central Dogma Revisited



DNA sequence

pre-mRNA

mRNA

Protein sequence

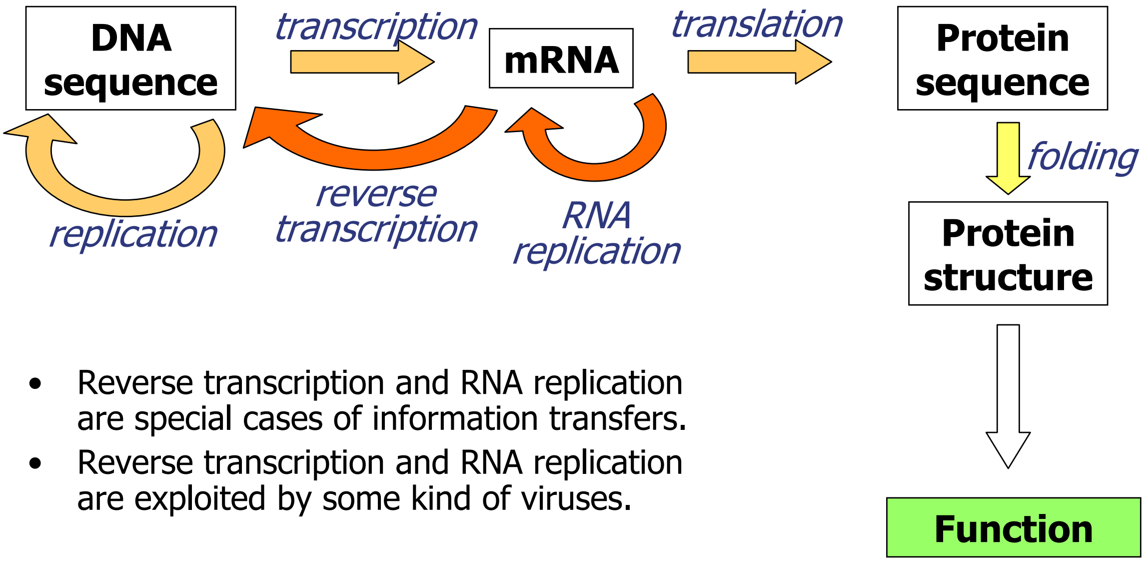
Precursor Protein

Protein structure

Function

- Splicing: removal of introns
- Post-translational modifications of a protein include many different possible processes
- Again: not so simple…

Central Dogma "Reversed"



DNA sequence

mRNA

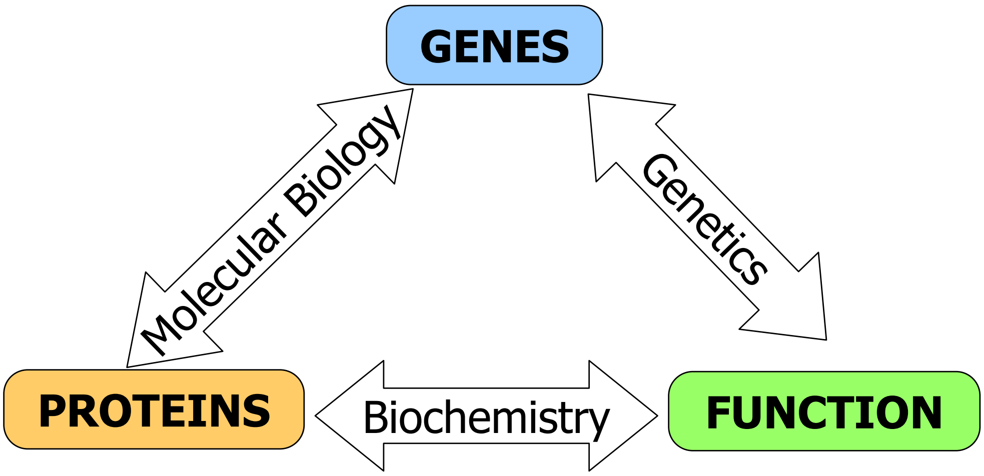
Protein sequence

Protein structure

Function

- Reverse transcription and RNA replication are special cases of information transfers.
- Reverse transcription and RNA replication are exploited by some kind of viruses.

Studying the Scenery



GENES

Molecular Biology

Genetics

PROTEINS

Biochemistry

FUNCTION

Information from the Sequence downwards…

- Information is coded into gene sequences
  - Sequence → Structure (Shape) → Function
  - Similar sequences → Similar Shapes/Functions?
  - Sequences are evolutionary related (phylogeny)

- Functions depends on movements/bindings of biomolecules
  - Correct function ↔ correctly folded proteins
  - Function can depend on multi-molecule complexes
  - The sequence-to-function paths are extremely complex

- Computational challenges
  - Sequence/structure similarities, interpretation of experimental data, folding/misfolding prediction, shape/function relation, analysis of molecular interactions and networks, etc.

What's Required?

- Reliable models
  - From Physics, Chemistry…
  - From Computer Science…
- Efficient algorithms
  - Depending upon model/process characteristics
- Exploitation of available information
  - Access to data banks, standard data format
  - Data from experiments
- A basic SW tool to work on models, algorithms, data: a scripting language

---

What Language?

Several languages are used
in the Computational Biology community

- Perl
  - Efficient in dealing with strings (also BioPerl toolkit)
- Python
  - Multiparadigm, flexible, good with sequences
  - Getting more popular (also BioPython modules)
- Ad-hoc languages
  - e.g. BioLisp dialect, Dynamite (for dynamic programming), etc.
- C++ required for speed
- Java (even a BioJava toolkit exists)
- Fortran (old routines)

Why Python?

- It’s a free, handy, efficient scripting language
- It can be used both standalone
  and as an integration glue for other sw tools
- A lot of specific libraries are available, such as

- SciPy
- NumPy
- BioPython

- …and it can be used in many other fields
  (not limited to Bioinformatics)!

Final Project

- Final test:
  presentation of the assigned project + oral discussion
- Project
  - Each project must be developed by X people
  - Each project must address a specific topic
    in Computational Biology, chosen with the teacher
  - Python + other tools (depending on the project goals)
    must be used
  - Projects are assigned 2/3 weeks before the class end.
- Final defense: Slide show + program(s) show +
  questions on class topics