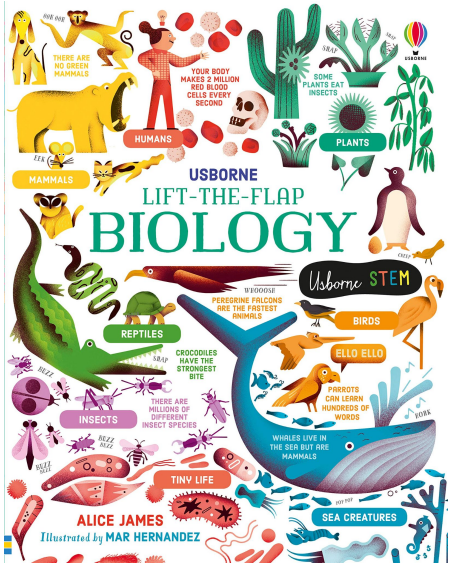
Foundations of AI's masterclass

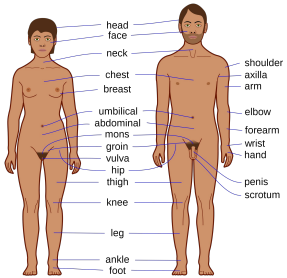
Structural Bioinformatics and learning

Biology

- The study of living systems
- Subfields :
  - Molecular biology, biochemistry
  - Biophysics
  - Marine biology
  - ...

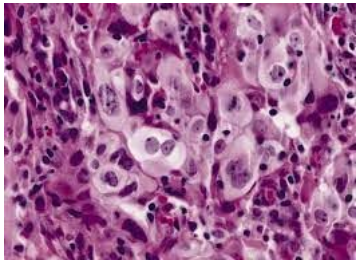


Fast molecular biology : zooming in !

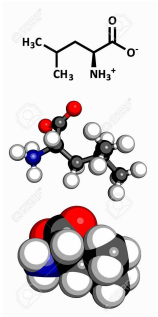


zoom

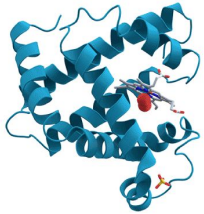
zoom



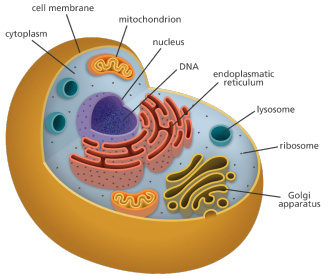
zoom



zoom

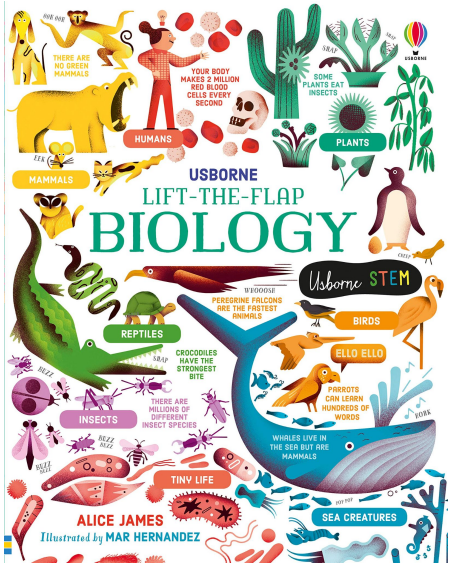


zoom



Some example applications

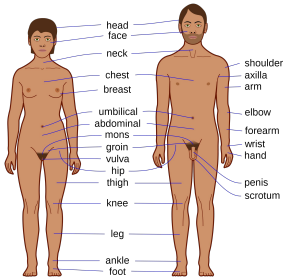
- Health : Covid vaccine
- Ecology : Understand ecosystems
- Energy : better batteries, solar energy
- ...

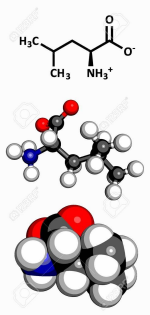


Incredible complexity

- We have not solved chemistry…
- … but biological processes are huge and intricate and very diverse

This has drawn this field to a lot of empirical approaches !





Problem solving in Biology

The classical steps of research in molecular biology :

- We have a biological system (an organ, an animal…) with open questions
- One of these questions could be of use, so we get money to investigate this question
- Try to isolate your system, perform measurements using cool techniques
- Analyze the data and draw a conclusion on the questions you were asking yourself

Specialized stem cell niche enables repetitive renewal of alligator teeth

Ping Wu [1], Xiaoshan Wu, Ting-Xin Jiang, Ruth M Elsey, Bradley L Temple, Stephen J Divers, Travis C Glenn, Kuo Yuan, Min-Huey Chen, Randall B Widelitz, Cheng-Ming Chuong

Affiliations + expand

PMID: 23671090 PMCID: PMC3670376 DOI: 10.1073/pnas.1213202110

Free PMC article

Hair cells, hearing and hopping: a field guide to hair cell physiology in the frog

M S Smotherman [1], P M Narins

Affiliations + expand

PMID: 10887064

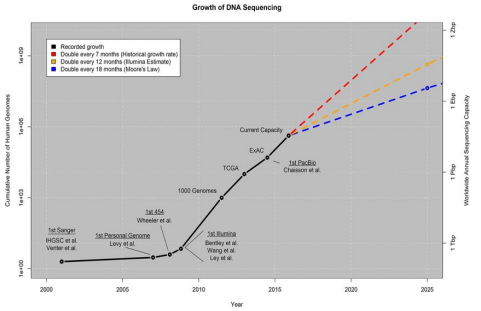
Abstract

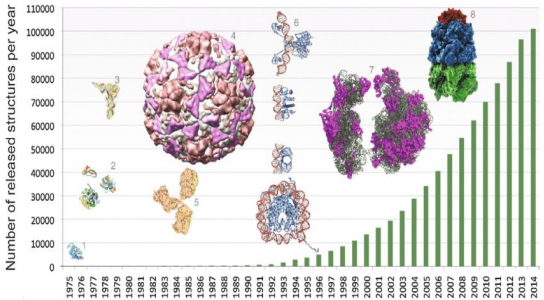
For more than four decades, hearing in frogs has been an important source of information for those interested in auditory neuroscience, neuroethology and the evolution of hearing. Individual features of the frog auditory system can be found represented in one or many of the other vertebrate classes, but collectively the frog inner ear represents a cornucopia of evolutionary experiments in acoustic signal processing. The mechano-sensitive hair cell, as the focal point of transduction, figures critically in the encoding of acoustic information in the afferent auditory nerve. In this review, we provide a short description of how auditory signals are encoded by the specialized anatomy and physiology of the frog inner ear and examine the role of hair cell physiology and its influence on the encoding of sound in the frog auditory nerve. We hope to demonstrate that acoustic signal processing in frogs may offer insights into the evolution and biology of hearing not only in amphibians but also in reptiles, birds and mammals, including man.

Biology and machine learning

- These steps involve gathering (a lot of) empirical data
- The resulting relationships are not always understood

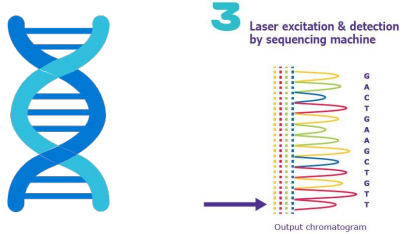
=> This looks like a mission for machine learning !



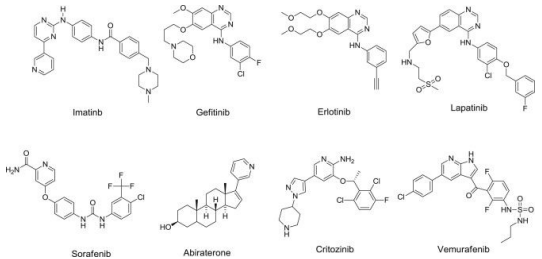


Biology and machine learning

- Genomic data :
  - Sequencing DNA
  - Predict cancer/not cancer



- Drug design
  - Get millions of small molecules
  - Predict toxic/not toxic



Challenges for this data

- Data is touchy
  - Privacy
  - Dangerous products

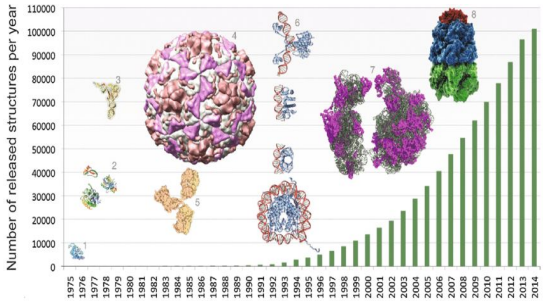
- Data is multimodal :
  - Genomics
  - Pictures
  - Medical file

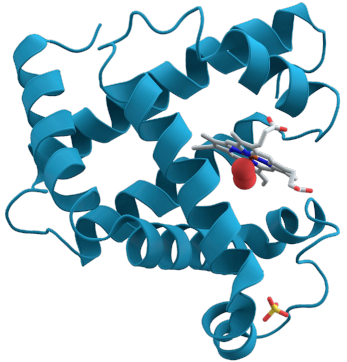
Vol de données : le profil atypique du « hacker » des hôpitaux ...

Oct 10, 2021 — Vol de données : le profil atypique du « **hacker** » des **hôpitaux** de Paris. Analyse. Un étudiant français de 22 ans a reconnu être l'auteur du vol ...

Structural Biology

- Living systems can be described as 3D molecular objects evolving through time
- Several techniques to obtain this data (crystals, NMR, cryo-EM…)
- Most of their properties can be explained using this representation (most complete description)

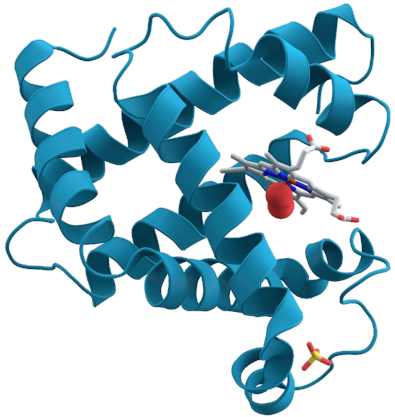




Chemical events

- Intricate chemistry happens between the biological molecules (very big, in complex, evolving solutions)...
  - Change of shape
  - **Binding**

- Chemically understanding these events is very hard

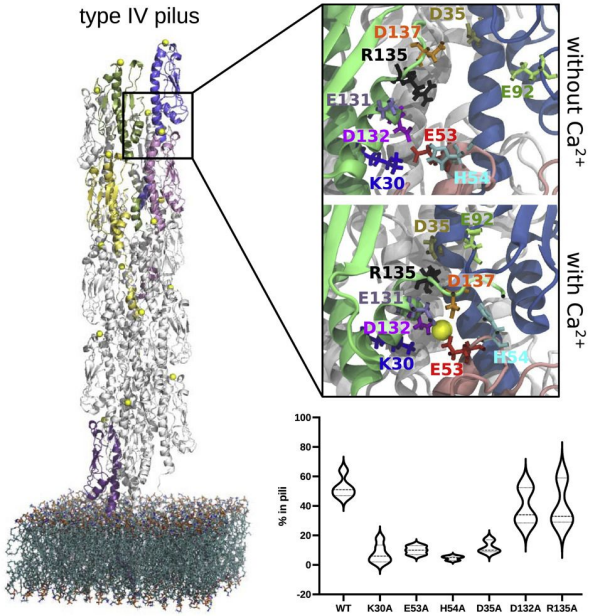


Structural Biology

- The goal of structural biology is to understand the chemical events thanks to the structures of the objects :
  - => Lock-key model

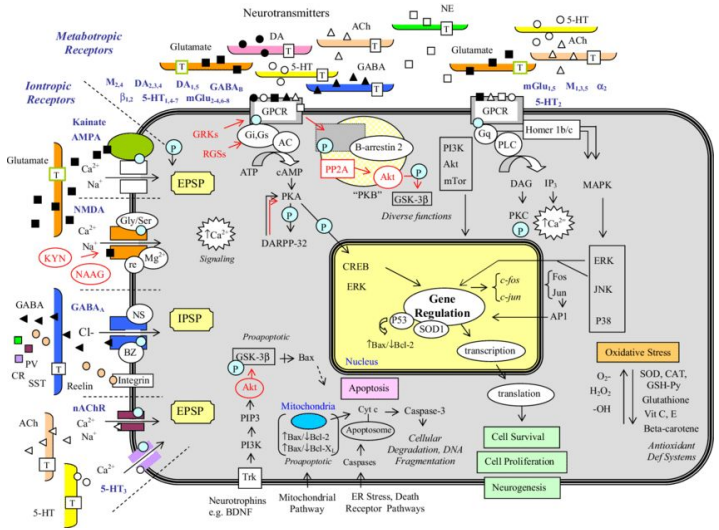
- There are two main goals :
  - Finding the structures
  - Using the structure to predict function

- The structures serve as high level models of the intricate chemistry to predict other properties



Pathways

- Chemical events are organized in cascading molecular pathways :
  - Something happens that changes something else…
  - … that in turn changes two things…
  - … that in turn change 10 things…
  - …



Takeaways and questions

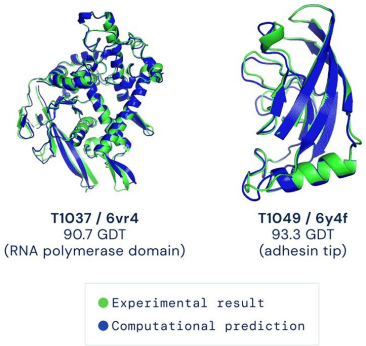
- Biology is this huge field of studying living systems, heavily relying on empirical data
- Structural biology is a subfield concerned with the shape of biological molecule, in turn explaining their function
- Theoretically, from this structure, we can predict any chemical event. Along with the molecular pathways, this would be an extensive understanding of biology

Structural Biology and machine learning

- Predicting the structure
- From the structure predict the **function**

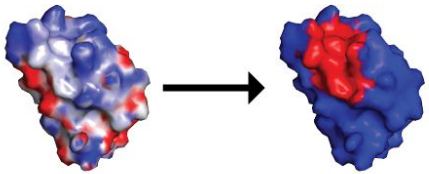
Predicting the structure

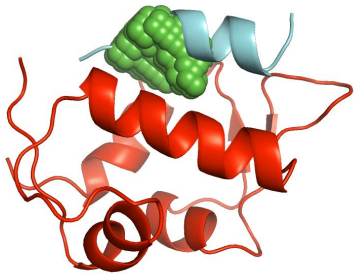
- This a long-standing problem (>60 years)
- The main landmarks were experimental techniques
- AlphaFold 1&2 was a significant milestone



Predict the function

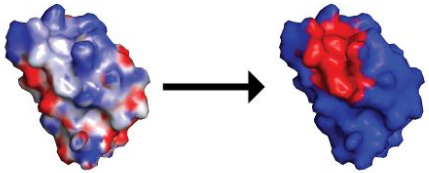
- Only one structure, but many functions !
- Prediction of binding :
  - DNA, RNA, Proteins
  - Small molecules (*target based*)
- Prediction of more abstract and general properties (QSAR)
  - Presence or absence of disease
  - Stability of a compound…





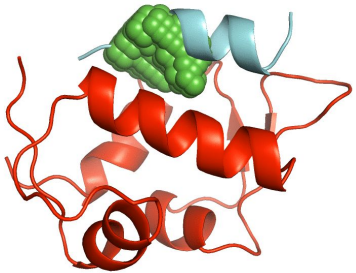
The main challenge

- The whole field revolves around the structure of these objects

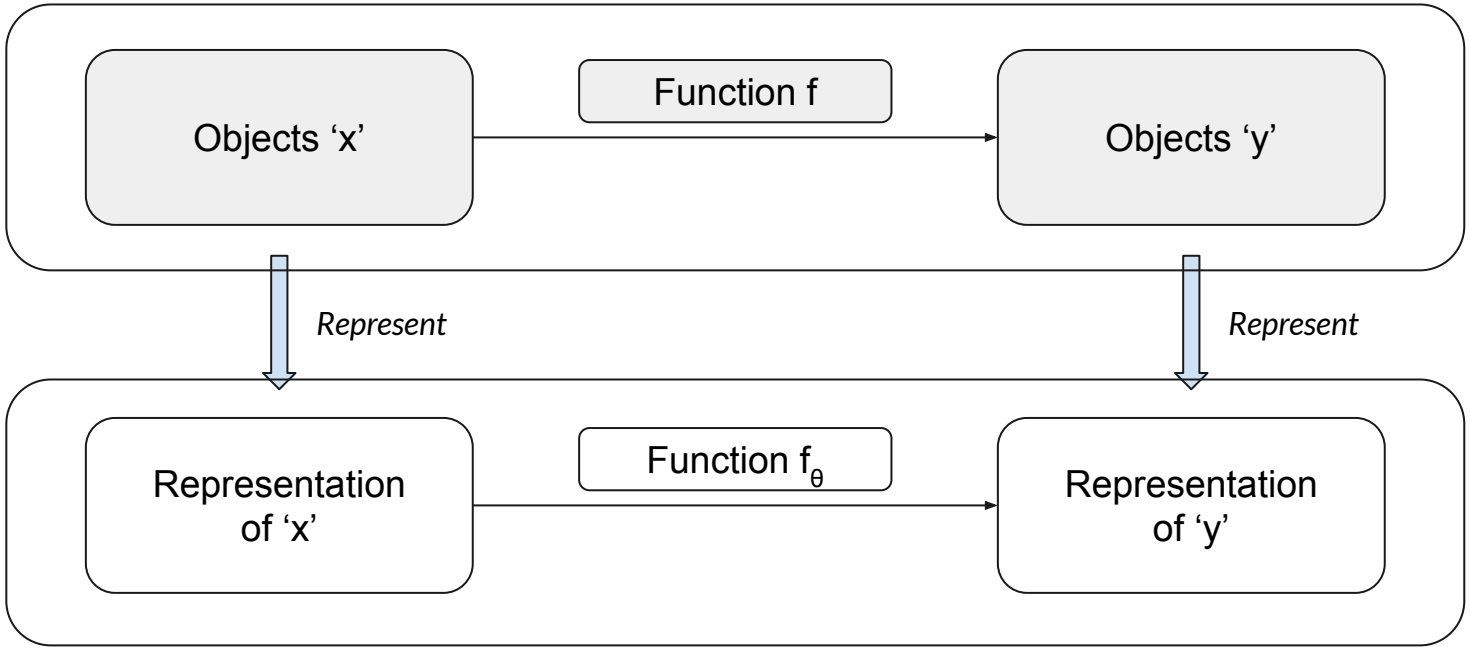


Can the machine learning algorithms take that into account ?

Remember an old slide…



Pre-Statistical ML

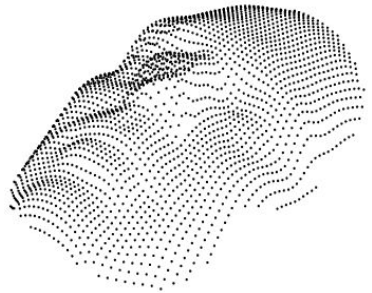


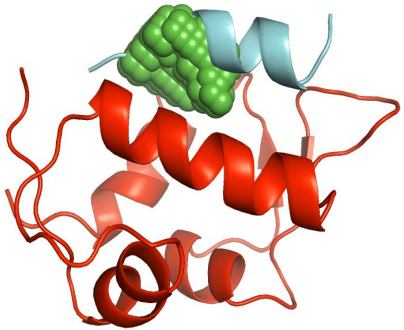
Hidden truth

Learnt relationship

Representing geometric objects

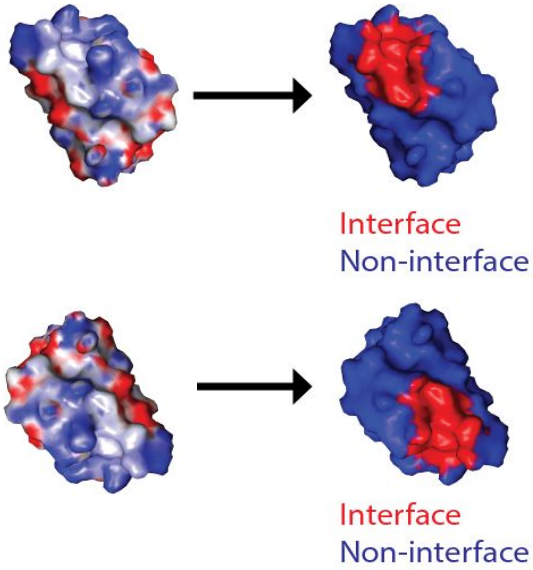
- The hidden truth is concerned with a structured concert
- How can we represent this complicated object into a computer ?
  - A photograph in 3D ?
  - Cloud of points ? (x,y,z coordinates)
- What is the best representation ?





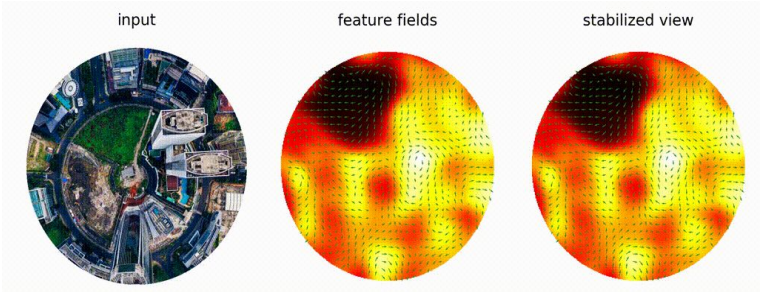
Rotation equivariance

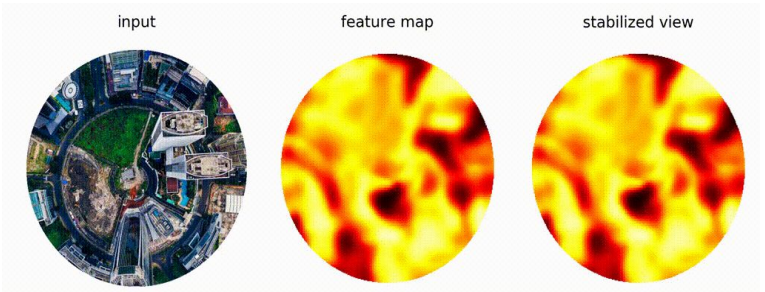
- An example problem : rotation invariance
- Gravity is not relevant for these objects
  - The nature of proteins does not depend on a top direction
  - Our hidden truth should not depend on it
- How do we represent something with no ‘up’ and ‘down’ in our computer ?



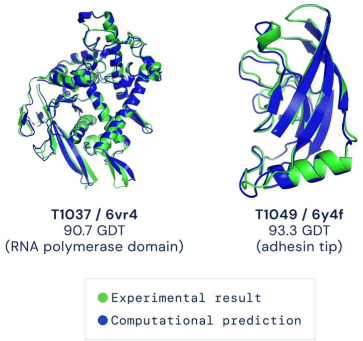
Geometric Deep Learning

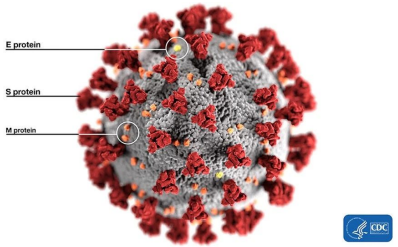
- There are dedicated representations and methods in machine learning
- The method is beyond the scope of the class





Example success stories





Some leads are being tested in the lab for potential Covid drugs

Takeaways and questions

- The structure aspect of structural biology makes it more challenging to learn from
- Some dedicated methods exist that help learning meaningful relationships
- This has lead to huge boosts and a complete paradigm shift in this field of research.

Thanks for your attention !