Introduction. Biophysics
(from a bioinformatics perspective)

Outline

- Definitions and concepts
  - Definition of molecular biophysics. Interaction with other subjects.
- Reference data. Experimental data and associated problems.
- Calculable magnitudes. Limitations
- Model Systems. Howto. Limitations and aproximations
- Validation and experimental design

Definitions

Karl Pearson 1892

- (from the Wikipedia): Biophysics or biological physics is an **interdisciplinary science** that applies the approaches and methods of physics to study biological systems. **Biophysics covers all scales of biological organization**, from molecular to organismic and populations. Biophysical research shares significant overlap with biochemistry, physical chemistry, nanotechnology, bioengineering, computational biology, biomechanics and systems biology.

Which scale?

- Molecular Scale: Molecular Biophysics.
  - Understand thermodynamic and kinetics aspects of the structure to function relationships in biomolecules. Overlaps with Biochemistry and Molecular Biology
  - We will focus on **Molecular Biophysics**
- Organism scale:
  - Animal locomotion, Biomechanics, Biomineralization, Motility. Overlaps with bioengineering, nanotechnology, systems biology, Medical biophysics
- Environment scale
  - Biogeophysics, Environment, Ecology

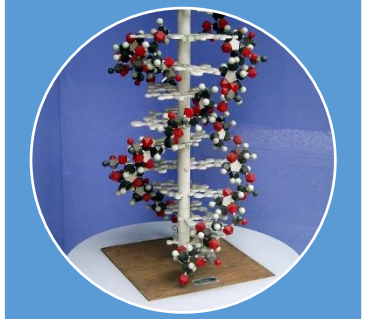
The Bioinformaticians' perspective

- Bioinformatics provides tools to help Biophysics
- Bioinformatics should understand:
  - **The nature of data**: Experimental data, structure data
  - **The principles underlying biophysics**: thermodynamics and kinetics
  - **Which biophysical magnitudes** can be calculated and how.
  - How to **build model systems** that can be analysed quantitatively
- Bioinformatics should
  - **Derive algorithms** to apply biophysics principles to data and models
  - Perform **data analysis** to draw conclusions usable to
    - Formulate hypothesis on the behaviour or biomolecules
    - Design experiments to validate such hypotheses

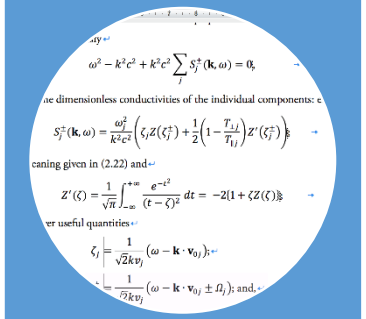
The process…

Understand the biological question

Recover available data



Define and build a model of the system under study



Identify magnitudes that can be calculated/ simulated

- Design algorithms

Fully understand the model quantitatively

- Draw and test hypothesis

Design validation experiments

1. Understanding the biological question (Not entirely your work …)

- Which is the question to answer
  - Does it require a "molecular" point of view?
  - Do we need a quantitative approach? Is it possible?

- Available experimental data
  - Biological systems, measured data
  - Do we really understand the nature of experimental data?
    - Ex. IC50 values are common in drug design experiments, but has no direct physical meaning. K_I should be used instead.
    - Ex. Intramitochondrial concentration of oxaloacetate a ca. 10^{-9} M. This gives 1 or 2 molecules!! per mitochondria

2. Recover available data

RCSB PDB
PROTEIN DATA BANK

ArrayExpress

ENA
European Nucleotide Archive

Model Archive

EGA
European Genome-phenome Archive

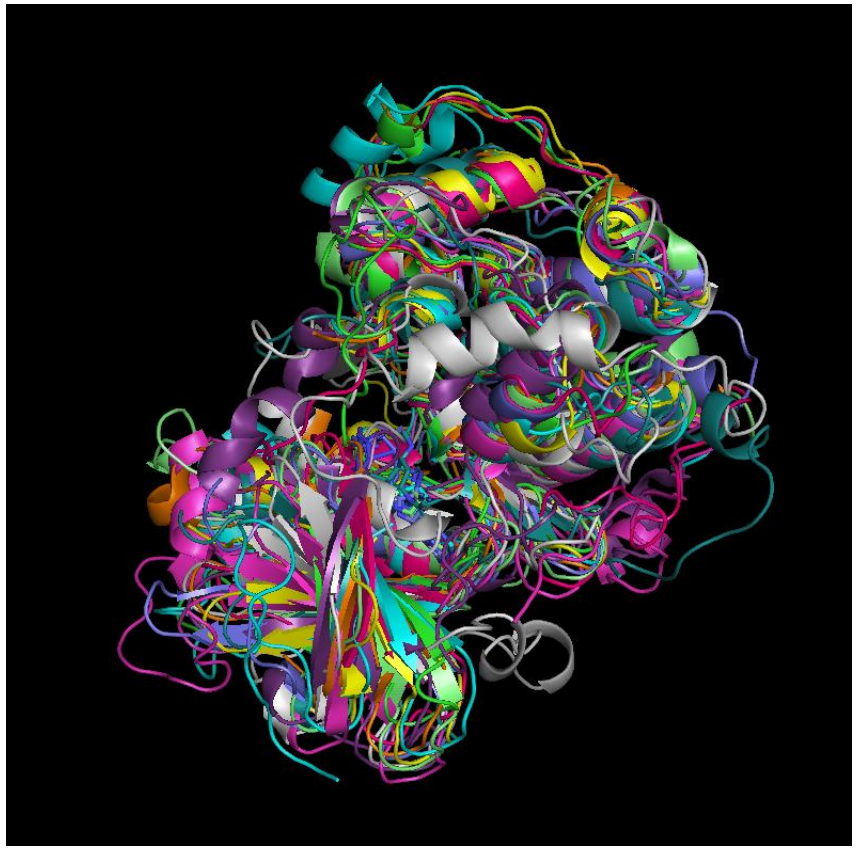
BioStudies

UniProt

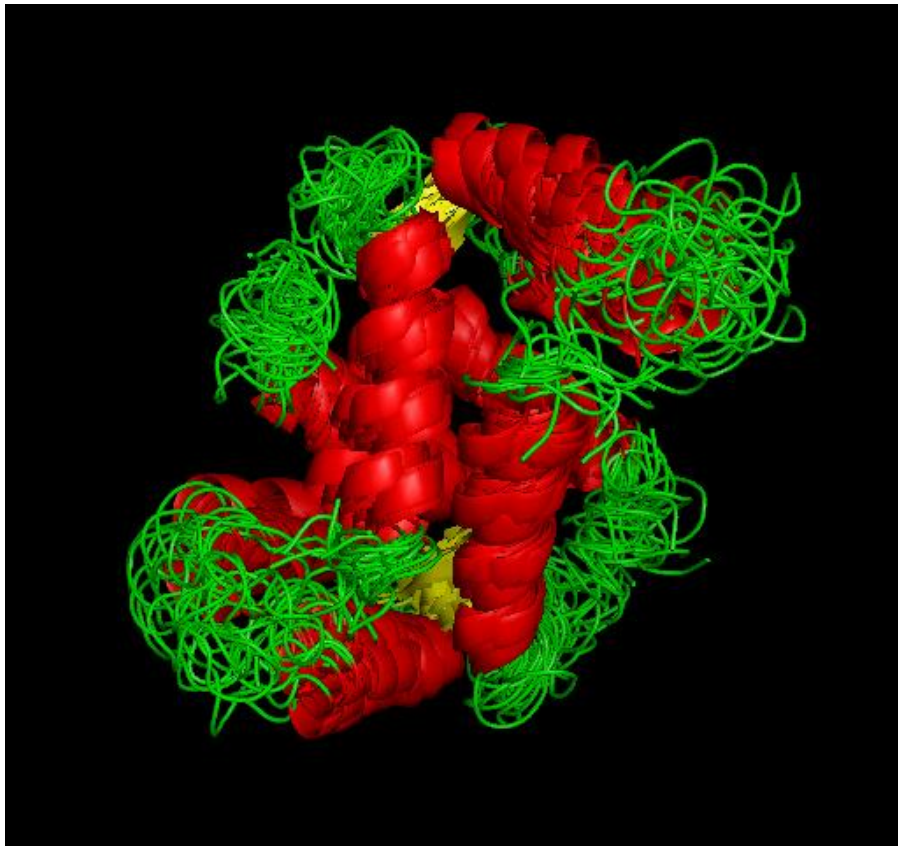
BioSamples

FAIR Data: Findable, Accessible, Interoperable, Re-Usable

Ex. Experimental protein structure ensembles



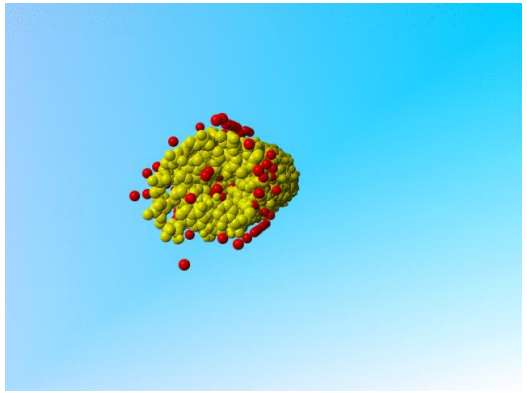
Xray: 1CM8 and other Prot. Kinases



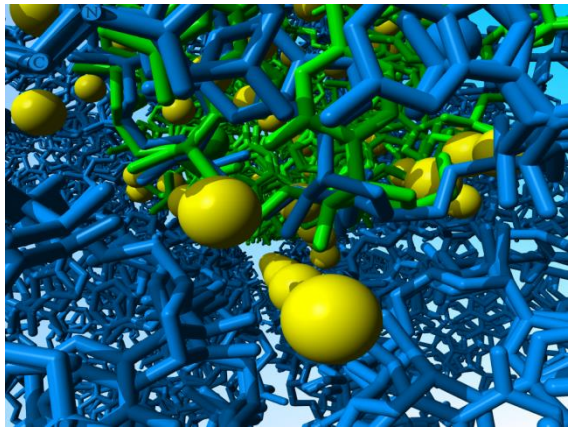
RMN: 1A03. Ca^{2+} Binding protein

The bad news about experimental data…

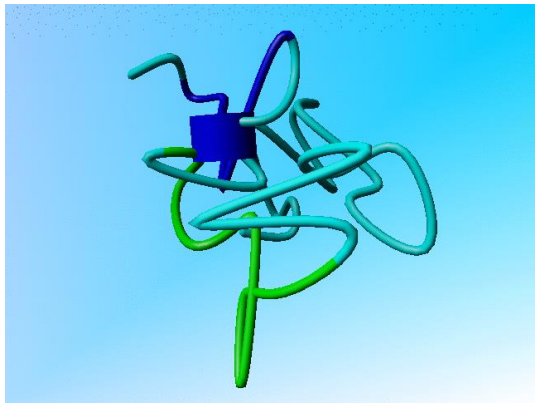
- may be just wrong
- may not be what we believe
  - Pretended “energies” (always ΔG?)
  - Unclear observables (e.g. IC50, arbitrary units)
- may have unknown dependencies
  - Experimental conditions may change the meaning of the data (e.g. crystallization pH)
- may not be accurate enough
  - We cannot be more accurate than the original data we use



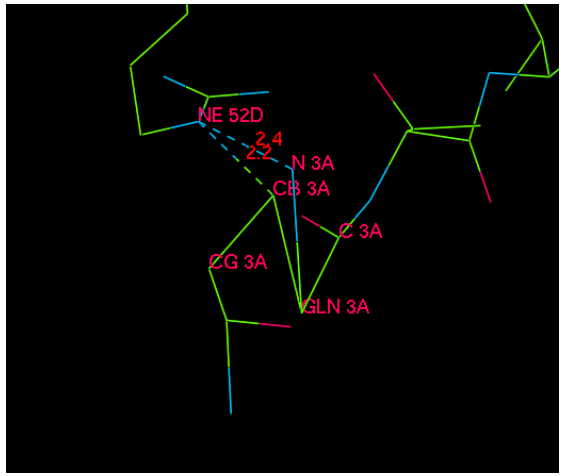
406d



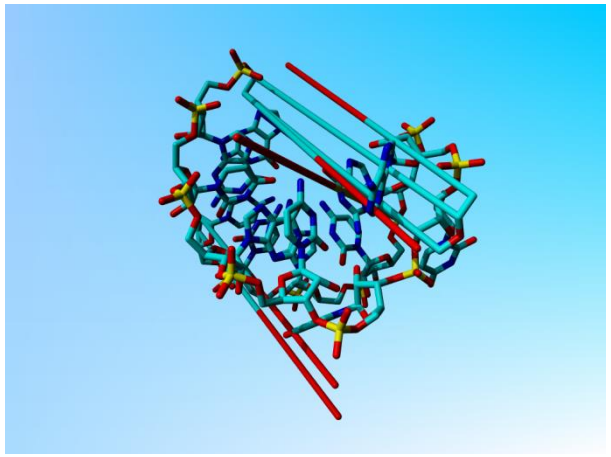
406d



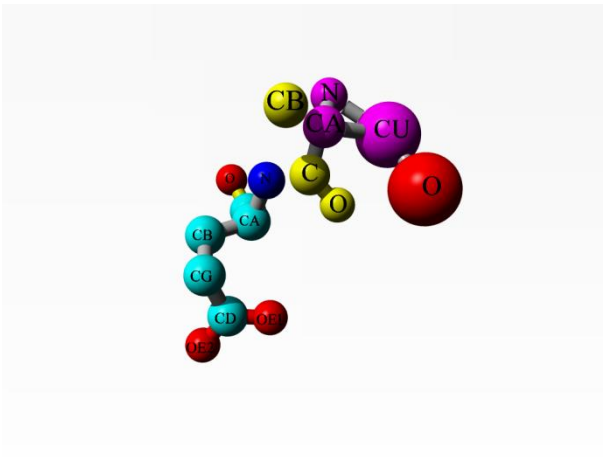
2PDE



1MN8



1l4c



1AG0

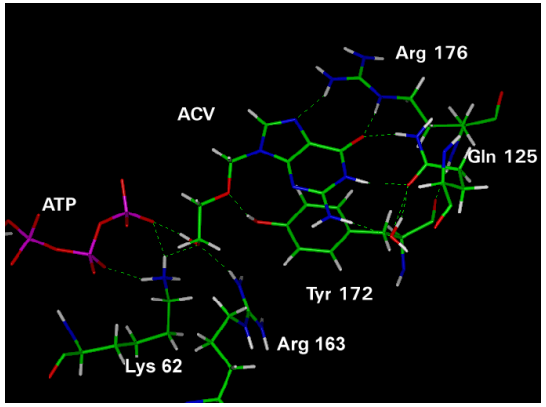
3. Building models

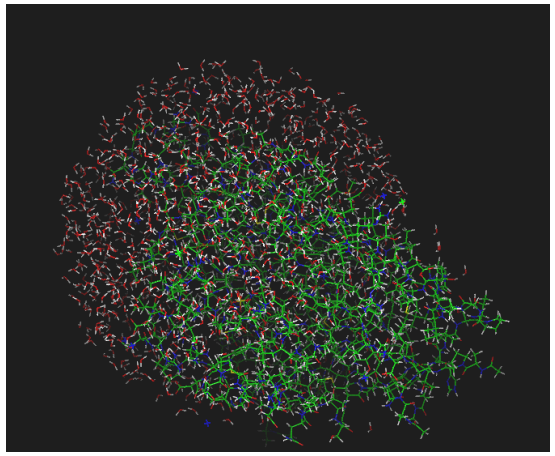
- Simplified representation of the biological (molecular) system (big and detailed enough to allow to answer the question)
- Relevant components should be included, but with too much detail the model can be unusable

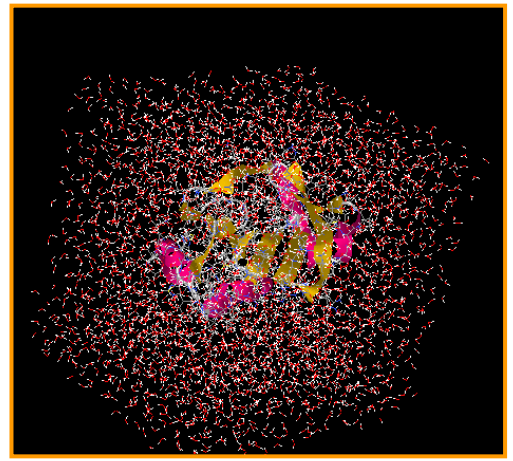
- Ex. Drug design
  - Protein active site (only aa that may interact with the drug)
  - Drug molecule(s)

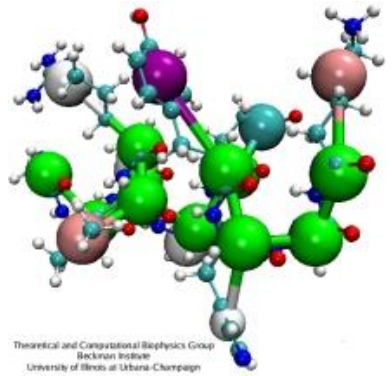
- Ex. Protein stability
  - Full protein molecule

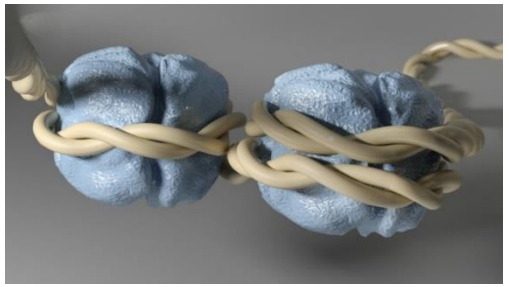
- Ex. Chromatin folding
  - Coarse grained model of chromatin fiber.











4. Discover calculable magnitudes and algorithms

- Biophysics is a "**quantitative**" science

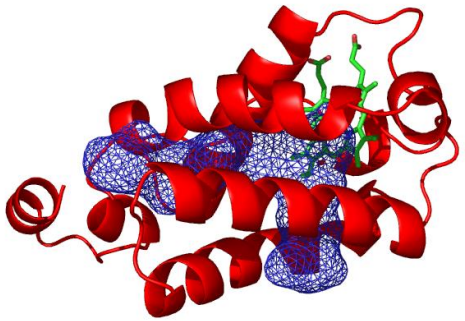
- Molecular geometry (distances, angles, surfaces, …)
- Energy (+ Enthalpy, Entropy, etc.)
- Binding energies
- Dynamics (conformation, fluctuations)
- Difussion rates
- Kinetic rates (reaction, transport, etc.)
- …

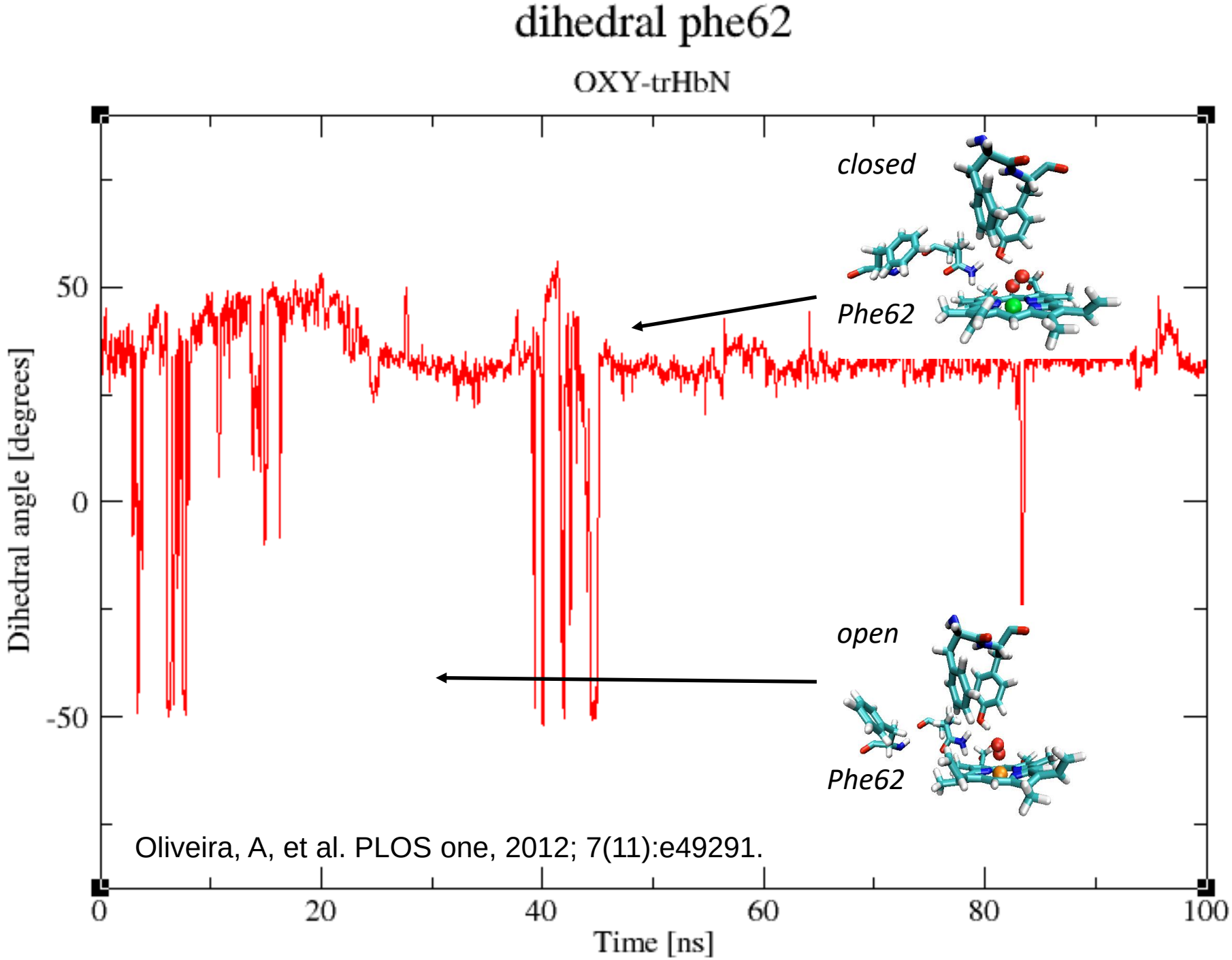
$$\Delta H(T) = \Delta H_m + \Delta C_P(T - T_m)$$

$$E_{vdw} = \epsilon \left(\left(\frac{R_{min}}{r} \right)^{12} - 2 \left(\frac{R_{min}}{r} \right)^{6} \right)$$

$$E_{elec} = \frac{1}{4\pi\varepsilon_o} \frac{q_i q_j}{\varepsilon_r r} = 332.16 \frac{q_i q_j}{\varepsilon_r r}$$

$$E = -RT \ln p$$

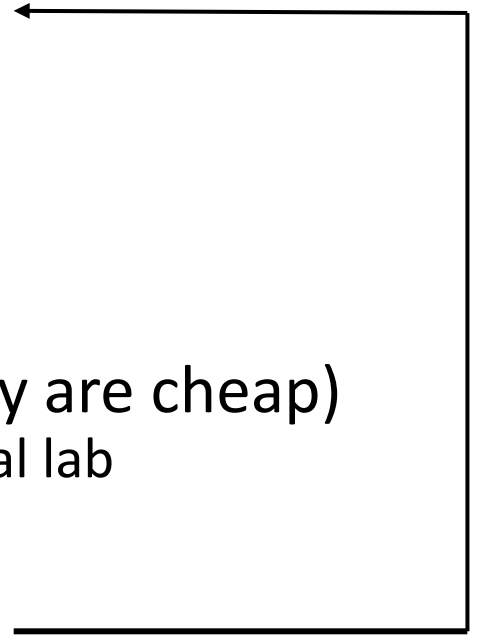




Oliveira, A, et al. PLOS one, 2012; 7(11):e49291.

5. Understand the model behaviour

- Translate the biological question to the model level
- Use available data to improve/check the model
  - Obtain parameters
  - Check data consistency
- Generate synthetic data to fill the gaps (with caution)
- Perform *in silico* experiments (as many as necessary, they are cheap)
  - Bioinformatics allows experiments that are not feasible in a real lab
- Redefine and improve the model as necessary

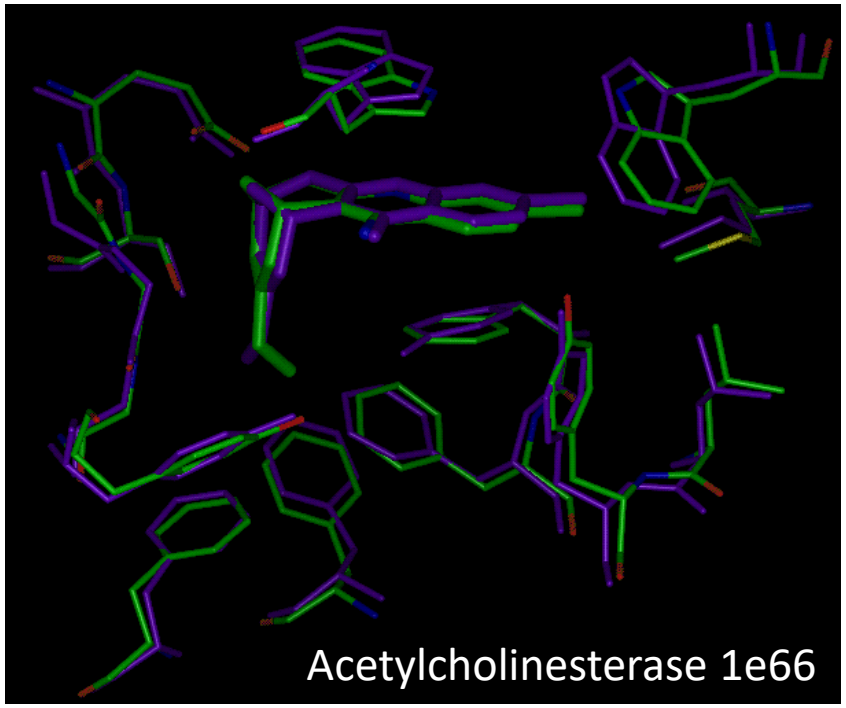


6. Validation

- *In silico* results help to understand the biophysics underlying the biological problem, but they are not the solution

- Experimental validation is required

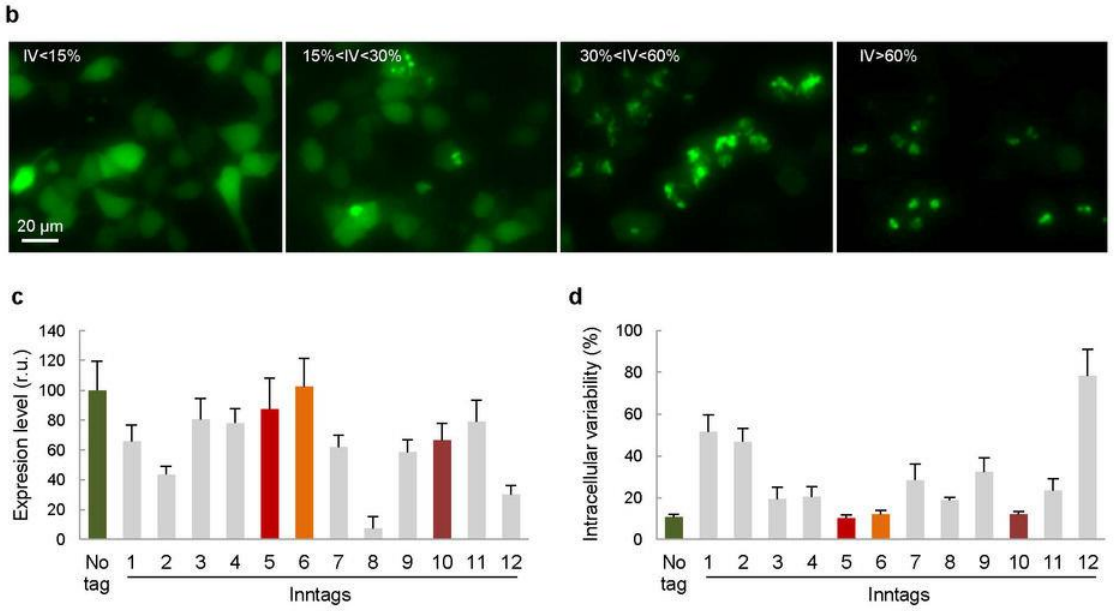
- *In silico* results helps to define which experimental measures should be done to test the hypothesis



X-ray Sussman and coworkers, Biochemistry 2002

a

tag	PDB id	Description	Organism	aa	Globularity
1	1BK8	Antimicrobial protein 1	Horse chestnut (*Aesculus hippocastanum*)	50	0.654
2	1BW3	Barwin, basic seed protein	Barley (*Hordeum vulgare*)	125	0.413
3	1ICX	Protein LLR18A	Yellow lupin (*Lupinus luteus*)	155	0.450
4	1PJW	Envelope protein	Encephalitis virus	111	0.389
5	1WHP	Allergen Phl p2	Timothy grass (*Phleum pratense*)	96	0.562
6	1WKX	Hevein isoform 2	Rubber tree (*Hevea brasiliensis*)	43	0.598
7	1X6R	Fimbrial protein	*Pseudomonas aeruginosa*	123	0.507
8	2JMH	Mite allergen Blo t5	Mite (*Blomia tropicalis*)	119	0.371
9	2JTY	Type-1 fimbrial protein, A chain	*Escherichia coli*	184	0.452
10	3BBG	Pollen allergen 5	Great ragweed (*Ambrosia trifida*)	40	0.587
11	3FT9	Allergen Phl p3	Timothy grass (*Phleum pratense*)	100	0.568
12	3K78	Major pollen allergen Bet v1-D/H	Birch (*Betula verrucosa*)	160	0.424



"Designed" Protein tags. Georgieva MV Nat Meth 2016

Summary

- Molecular Biophysics is about the quantitative understanding of the molecular aspects of Biology
  - Bioinformatics provides tools and algorithms to help biophysics
- Only biological problems that have a molecular component
- Model systems should be accurately built to include relevant components with the appropiate detail
- Model systems should be improved to be *consistent with* data available
- *In silico* experiments allow to fully understand the behavior of the model
- Experimental validation