

Selected topics in structural and computational biology

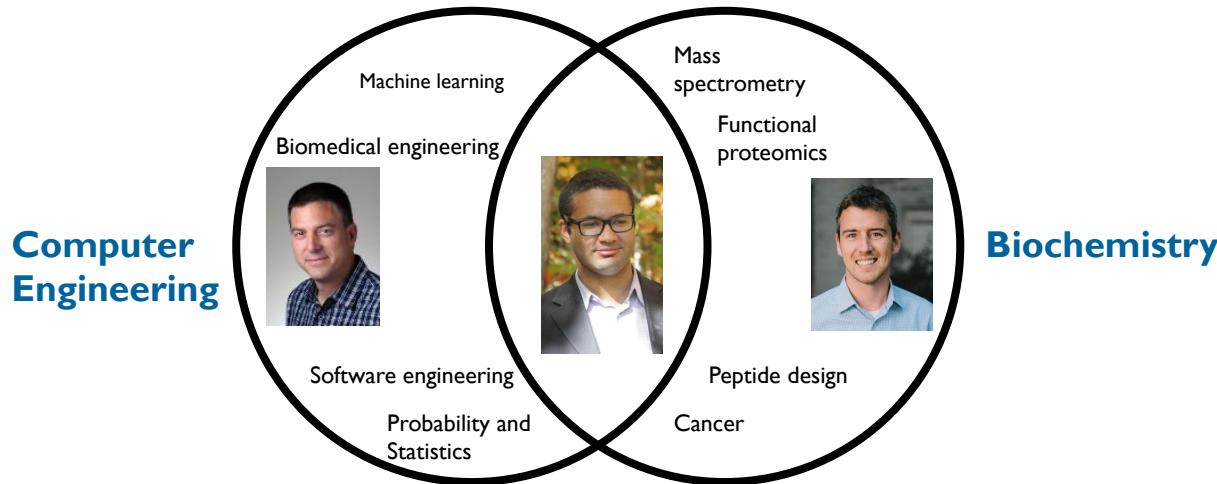
BIOC3202 - Guest lecture

Presented by
François Charih



Introduction

About me...



- Ph.D. student in Electrical and Computer Engineering (Carleton, circa 2023)
- M.A.Sc. in Electrical and Computer Engineering (Carleton, 2018)
- B.A.Sc. in Chemical Engineering (Ottawa, 2016)
- B.Sc. in Biochemistry (Ottawa, 2016)

Research interests: Computational biology, machine learning, *in silico* drug design, cancer

Hobbies: Reading, songwriting, and drinking coffee (lots!)

My first structure

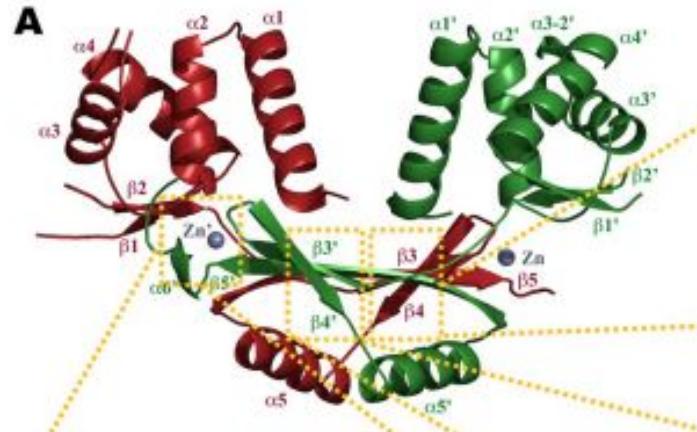
SCIENTIFIC REPORTS

OPEN

Functional insights into the interplay between DNA interaction and metal coordination in ferric uptake regulators

Received: 5 January 2018
Accepted: 11 April 2018
Published online: 08 May 2018

Sabina Sarvan¹, François Charih¹, Momen Askoura^{1,2}, James Butcher¹, Joseph S. Brunzelle³, Alain Stintzi¹ & Jean-François Couture¹



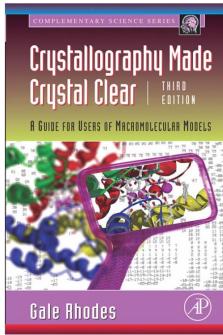
Lecture overview

I. X-ray crystallography (1h)

- a. Basic concepts
- b. Growing a crystal
- c. Acquiring a diffraction pattern
- d. Solving the structure
- e. Model assessment
- f. Advantages and shortcomings

2. Computational biology topics (45min)

- a. Computational prediction of protein-protein interactions
- b. *In silico* drug design
- c. Computational prediction of PTMs



Significant portions of this lecture build upon material from

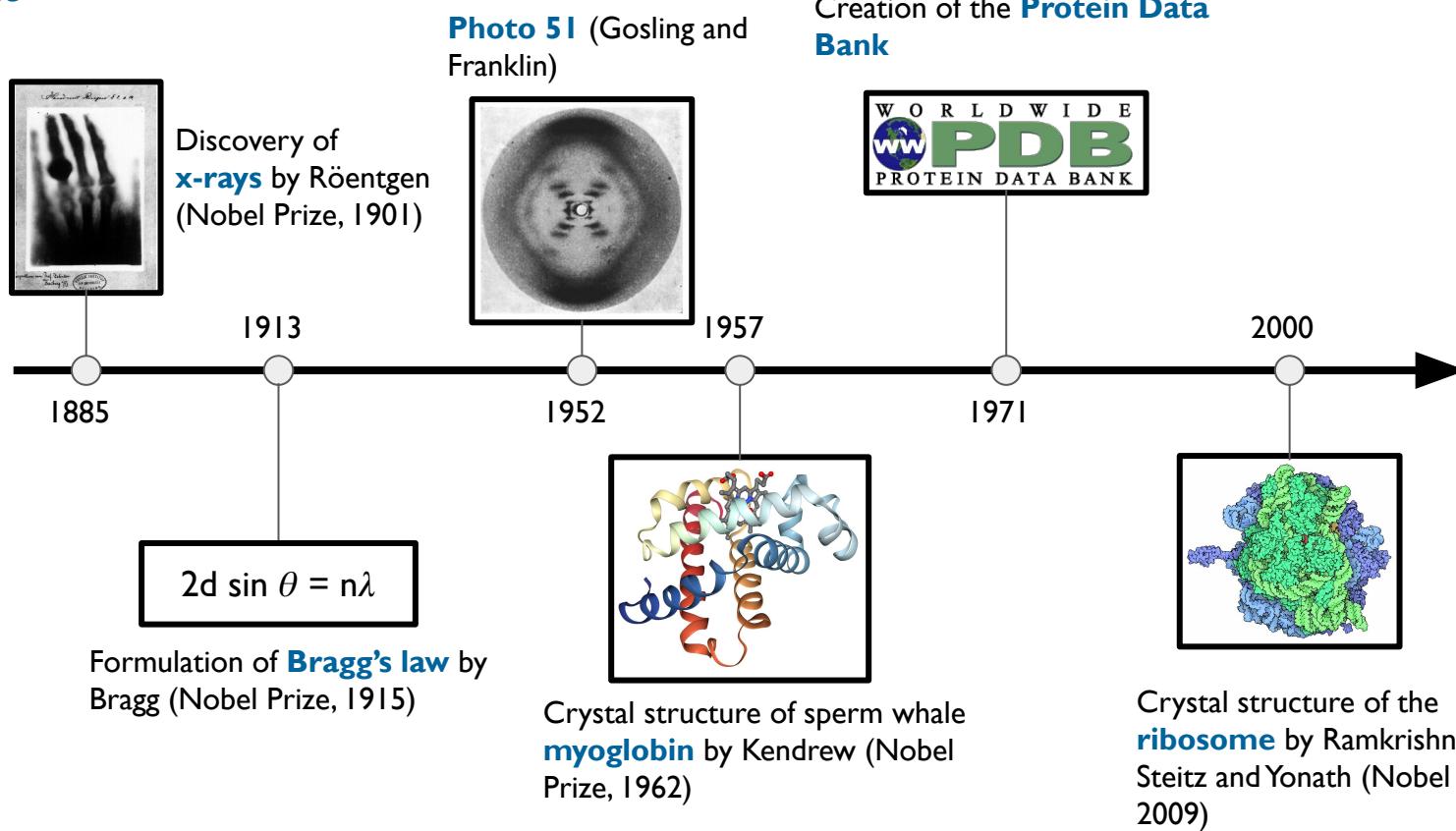
Rhodes, G. (2010). Crystallography made crystal clear: A guide for users of macromolecular models (Elsevier Science).

Part I - X-ray crystallography

Disclaimer: There is much more to be said about x-ray crystallography than can be said in 1h. I will focus on the most essential aspects.

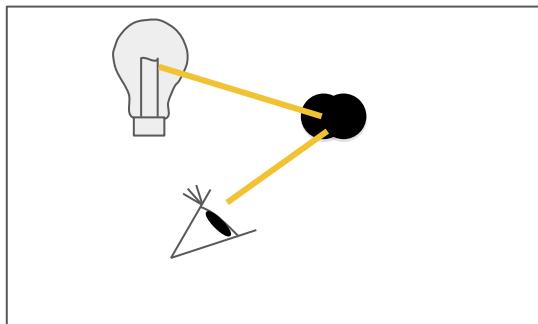
X-ray crystallography

Timeline



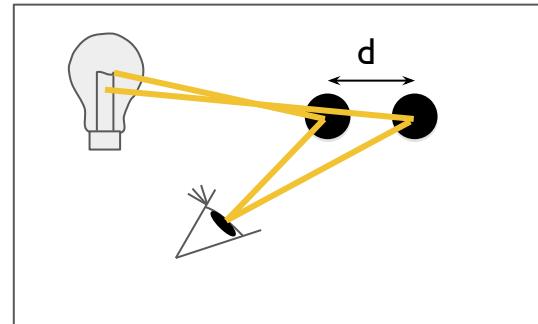
X-ray crystallography

Let's start with a definition



Situation 1

The observer sees one object, although there are two.



Situation 2:

The observer distinguishes two objects; distance d is large enough.

Resolution

The smallest distance that allows one to distinguish two distinct objects or points.

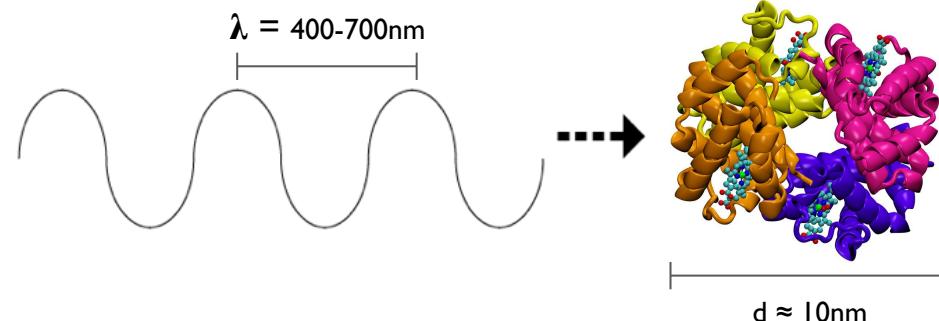
X-ray crystallography

The problem with visible light

General rule

The electromagnetic radiation must have a wavelength (λ) no larger than the object for the object to diffract incident radiation.

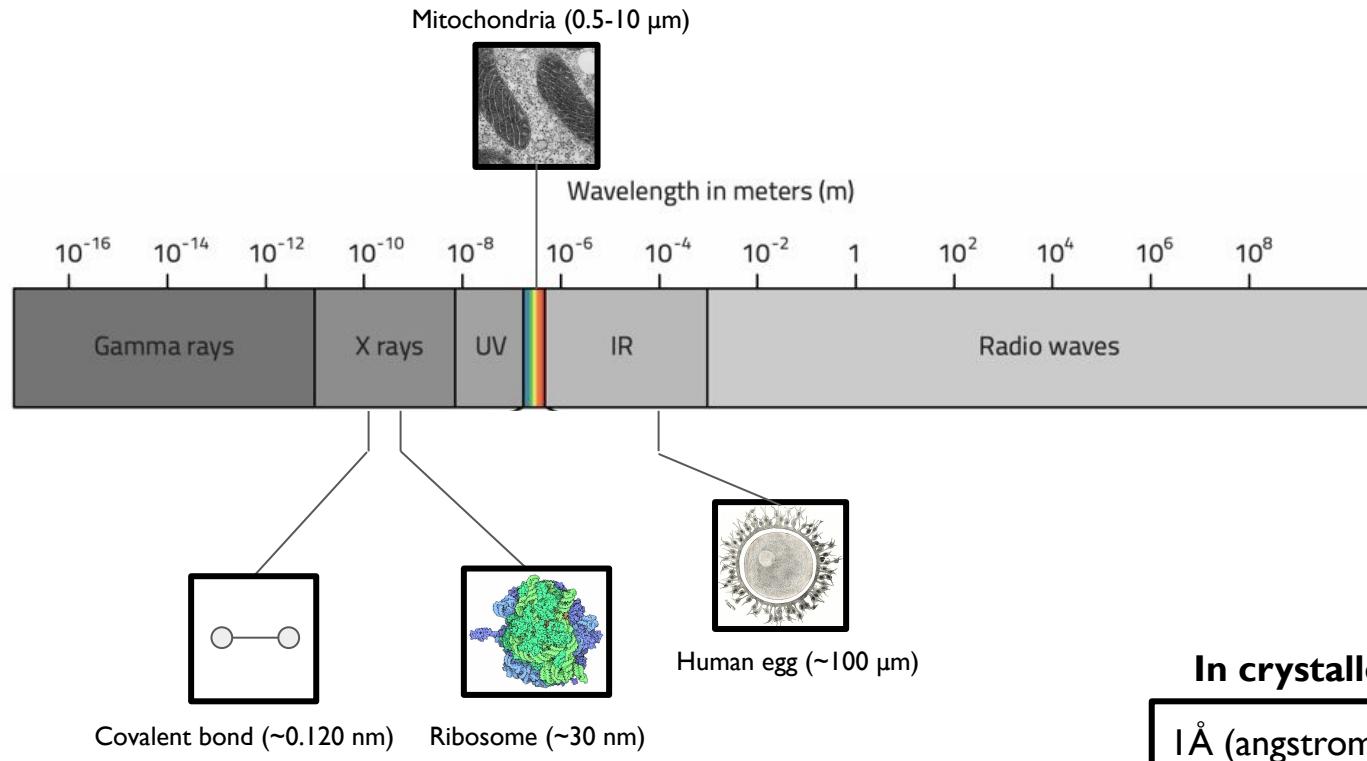
Problem



Resolution of visible light is too low to distinguish individual amino acids, and even less atoms.

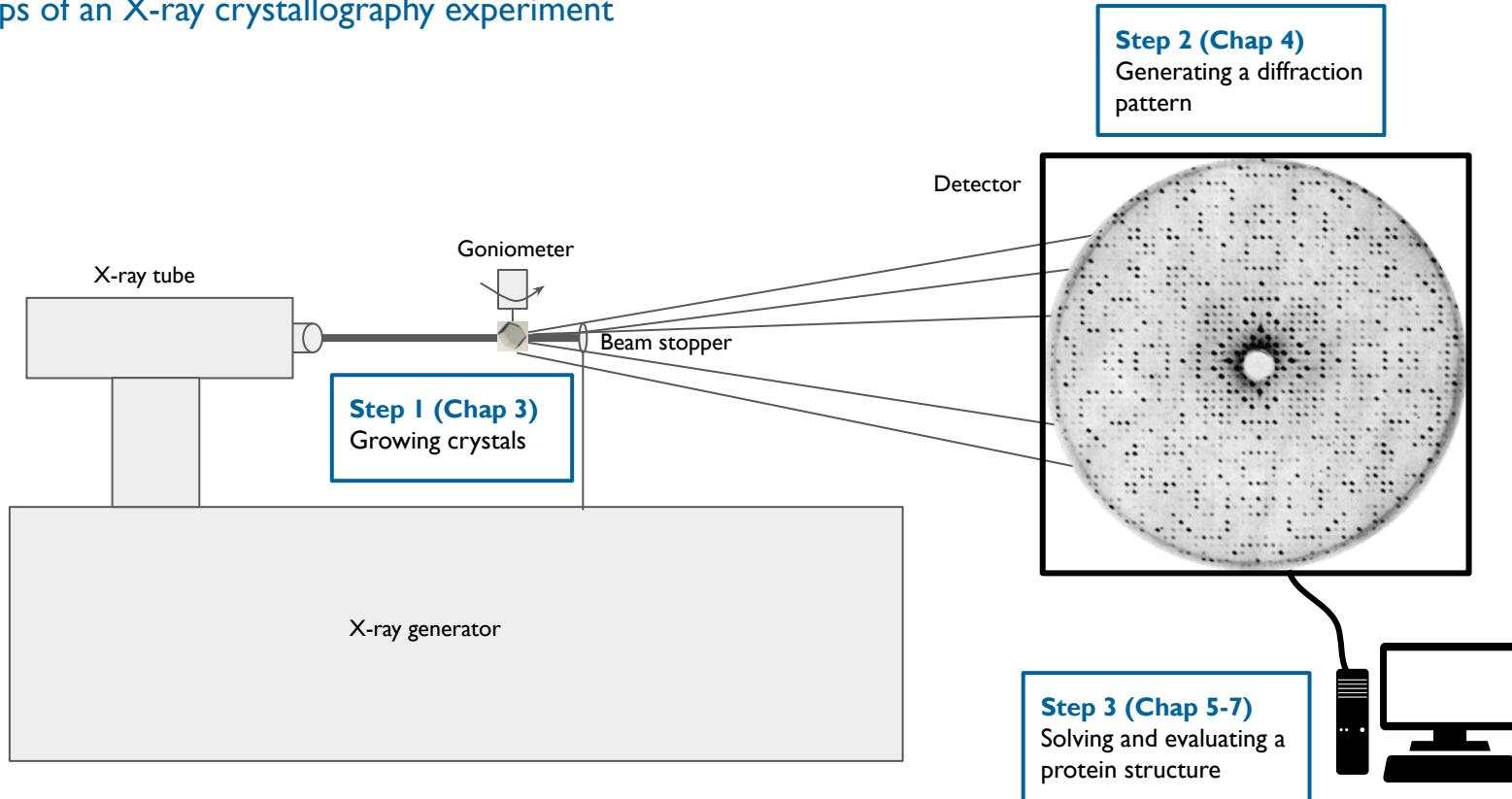
X-ray crystallography

X-rays unlock resolutions on the atomic and protein scales



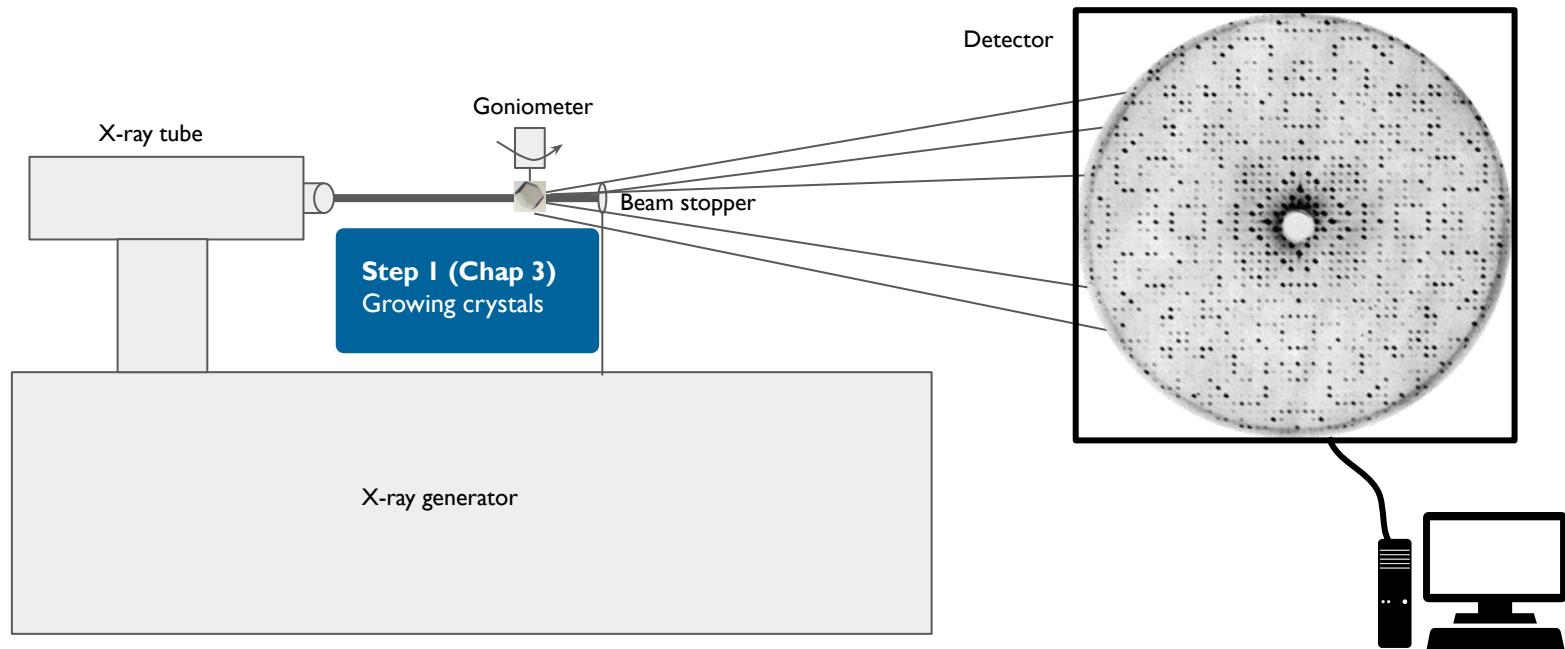
X-ray crystallography

The principal steps of an X-ray crystallography experiment



X-ray crystallography

The principal steps of an X-ray crystallography experiment



X-ray crystallography

Protein crystals: definitions

Protein crystal

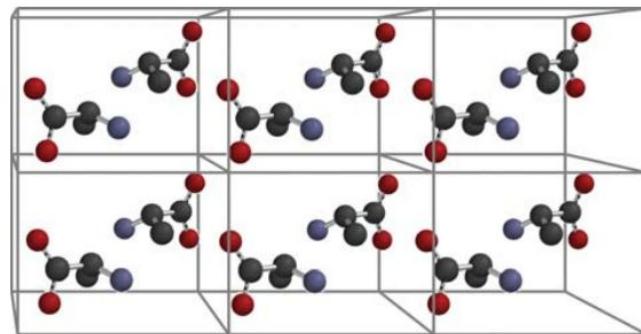
Orderly 3D array of proteins joined by non-covalent interactions (H-bonds)

Unit cell

Smallest repeating unit having the full symmetry of the crystal structure

Crystal lattice

Network of points formed by the corners of repetitive arrangement of unit cells

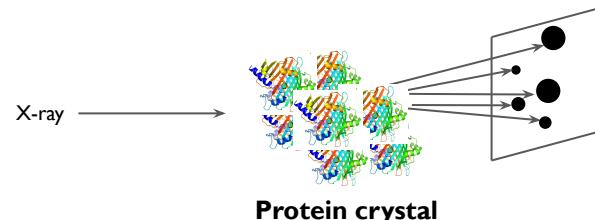
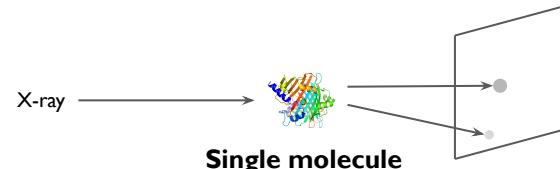


X-ray crystallography

Why grow crystals?

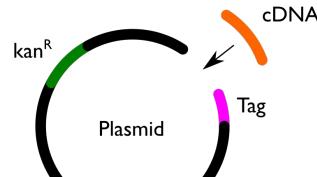
We grow protein crystals for multiple reasons:

- 1) Single proteins are **weak scatterers**
 - Weak reflections
- 2) X-ray exposure is **destructive**
 - Protein rapidly degraded due to generation of free radicals
- 3) Crystals are a **uniform** assembly of proteins with nearly identical orientations
 - Coherent X-ray scattering, unlike solution

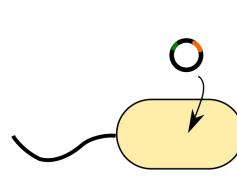


X-ray crystallography

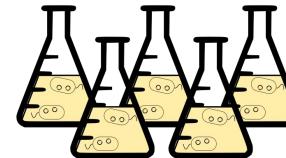
Growing crystals requires large quantities of pure protein



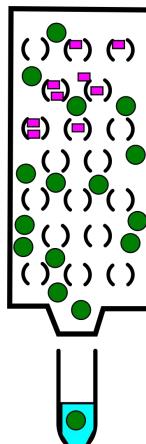
Cloning



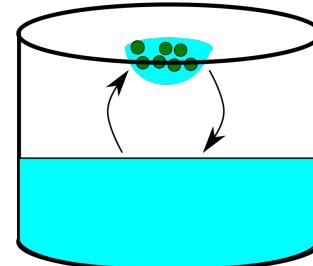
Transformation



Overexpression



Purification



Crystallization

X-ray crystallography

Protein crystallization

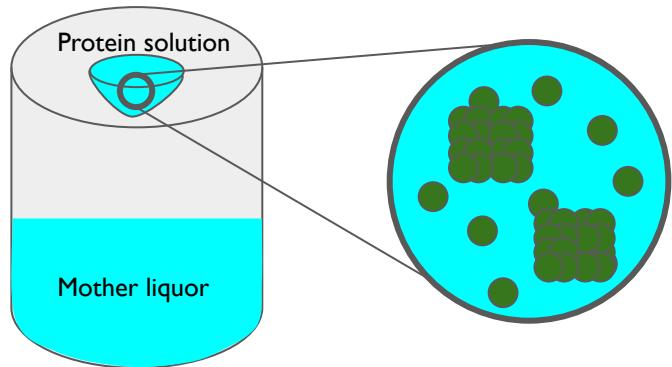
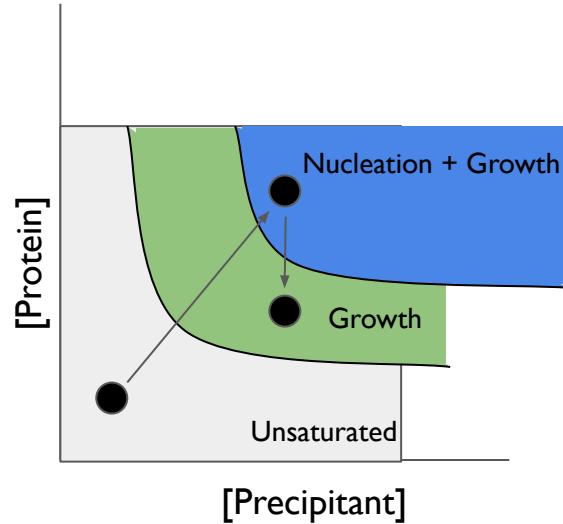
- Crystal formation requires **supersaturation**
 - i.e. exceed solubility threshold
- Two phases:
 - **Nucleation:** formation of molecular clusters that initiate growth
 - **Growth:** organized deposition of protein molecules around the nucleation point
- Impurities hamper nucleation
 - High purity is essential
- Control of **saturation speed** is pivotal
 - Slow precipitation => large crystals
 - Rapid precipitation => small crystals or amorphous solids

X-ray crystallography

Crystallization experiment - vapour diffusion method

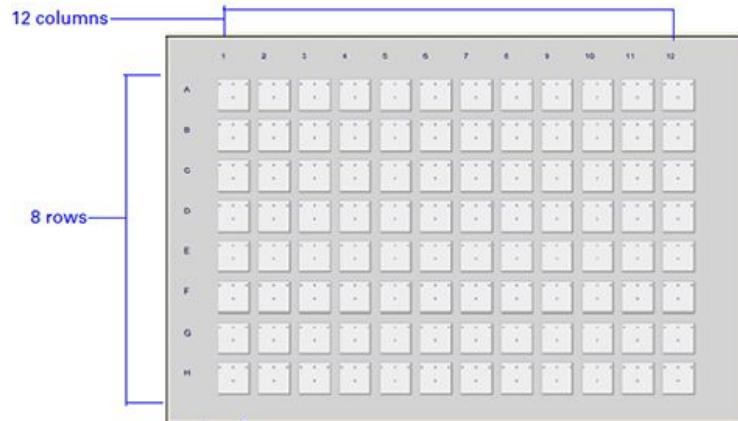
I. Unsaturated phase

- Low protein concentration



X-ray crystallography

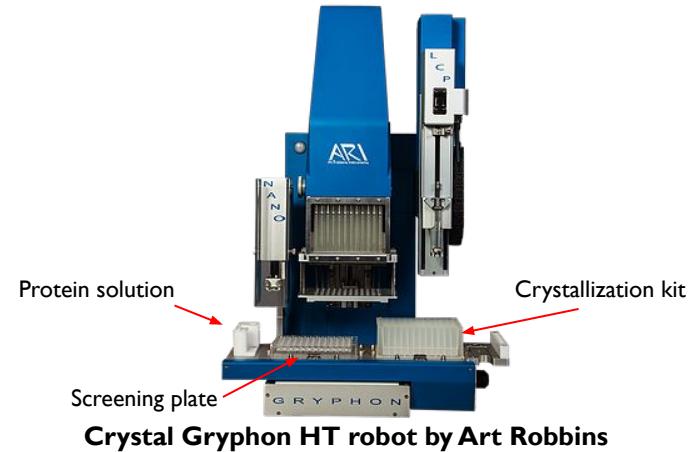
Crystallization experiment - high-throughput screening



A crystallography HT screening plate

Typical experiment requires loads of protein

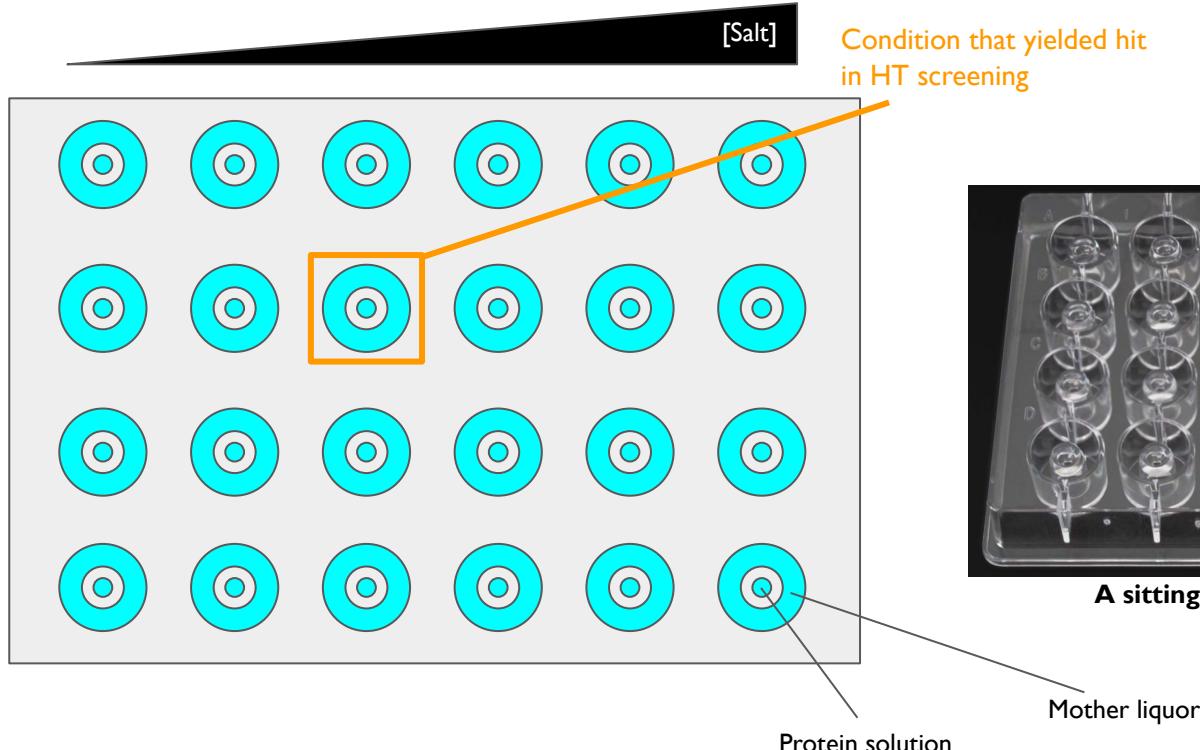
$$96 \text{ wells} \times (1 \mu\text{g} + 2 \mu\text{g} + 3 \mu\text{g}) \approx 0.6 \text{ mg}$$



A HT crystallization screening kit

X-ray crystallography

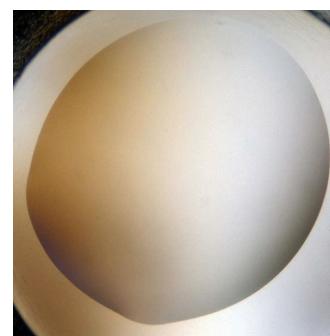
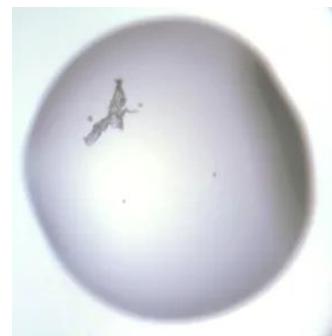
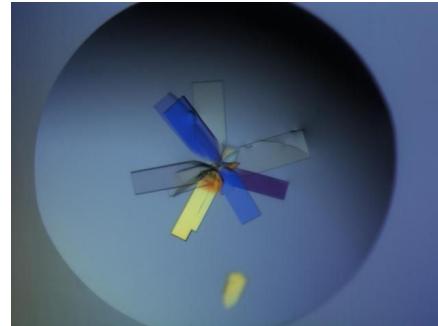
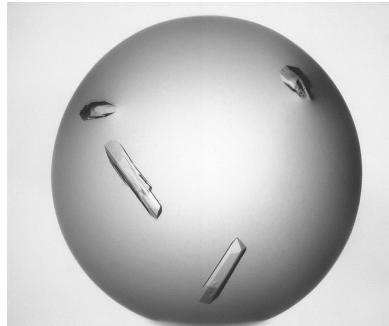
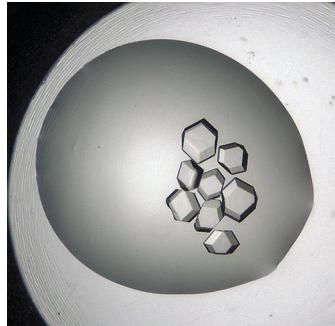
Crystallization experiment - optimizing hits



A sitting drop optimization plate

X-ray crystallography

When you look at your plate under the microscope...



Amorphous precipitate

Junk

Clear drop



X-ray crystallography

Sometimes a protein will not crystallize

Sometimes, proteins will not crystallize or yield inadequate crystals... what can we do?

Approach 1: Generating mutants

- Mutating disordered residues on the surface to low entropy residues
- Thermodynamic advantage: increase in $\Delta S_{\text{crystallization}}$
- Facilitate crystal packing with smaller, ordered residues

Approach 2: Seeding

- Use a small crystal as a seed around which a protein can grow

Approach 3: Stabilization with a ligand

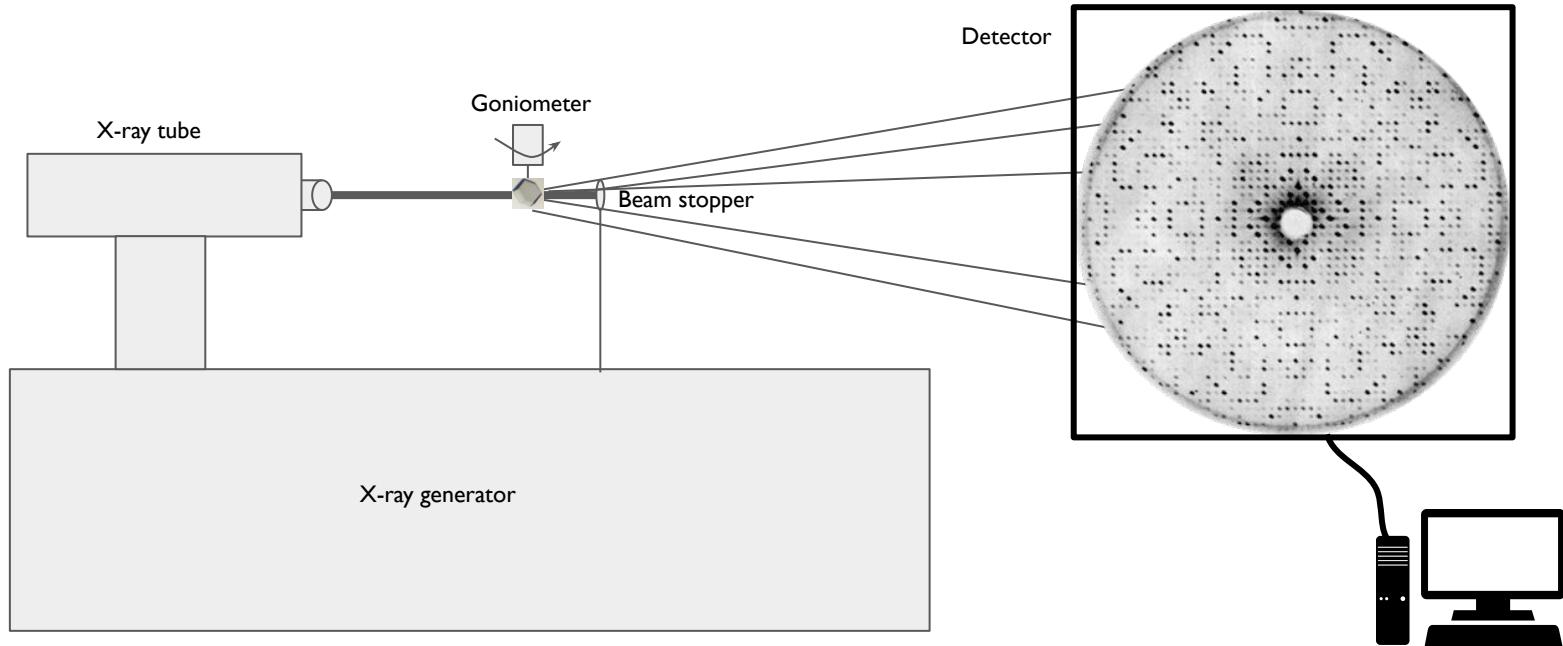
- Ligands (DNA, co-factor, peptide, etc.)
- Increase in protein rigidity due to interactions with ligand
- Conformational changes that facilitate crystal packing

X-ray crystallography

The principal steps of an X-ray crystallography experiment

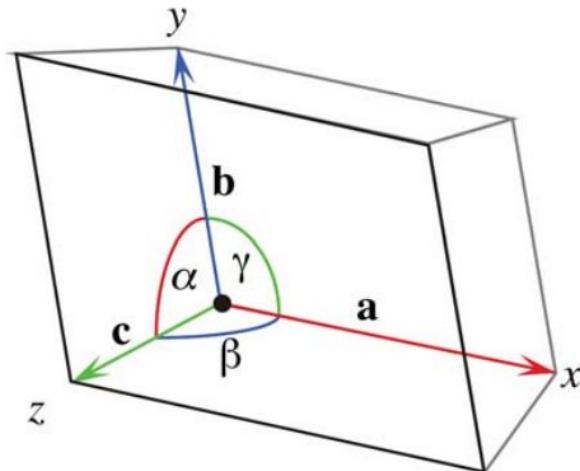
Step 2 (Chap 4)

Generating a diffraction pattern



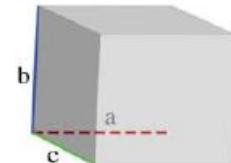
X-ray crystallography

Unit cell - defining a coordinate system



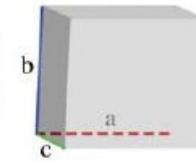
Cubic

$$a=b=c, \alpha=\beta=\gamma=90^\circ$$



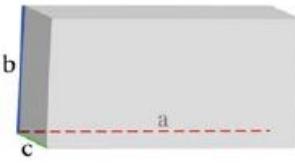
Tetragonal

$$a=b \neq c, \alpha=\beta=\gamma=90^\circ$$



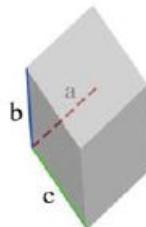
Orthorhombic

$$a \neq b \neq c, \alpha=\beta=\gamma=90^\circ$$



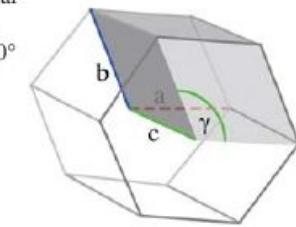
Rhombohedral

$$a=b=c, \alpha=\beta=\gamma \neq 90^\circ$$



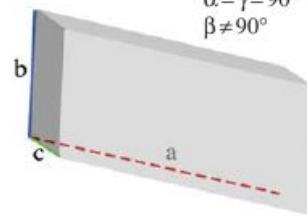
Hexagonal

$$a=b=c, \alpha=\beta=90^\circ, \gamma=120^\circ$$



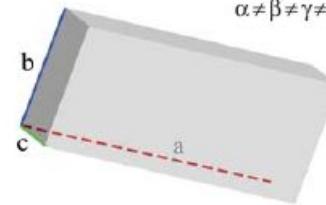
Monoclinic

$$a \neq b \neq c, \alpha=\gamma=90^\circ, \beta \neq 90^\circ$$



Triclinic

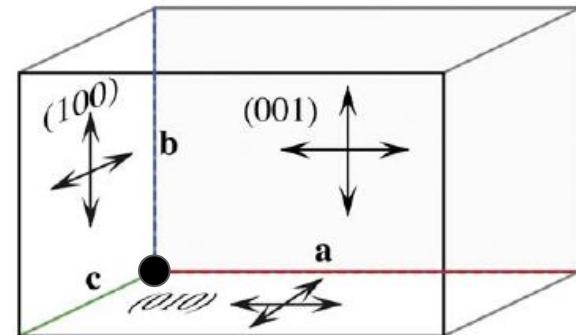
$$a \neq b \neq c, \alpha \neq \beta \neq \gamma \neq 90^\circ$$



X-ray crystallography

Miller indices

- Reflections on the detector are X-rays reflected by planes of electron density in the crystal
- The Miller indices hjk are used to refer to a plane of reflection within the unit cell
 - Integers
 - Sign defines orientation of plane
- The indices indicate the **number of planes that intersect the corresponding axis** in one unit cell
- Can indicate a plane or group of plane
- Together, can represent all the possible planes of reflection in unit cell

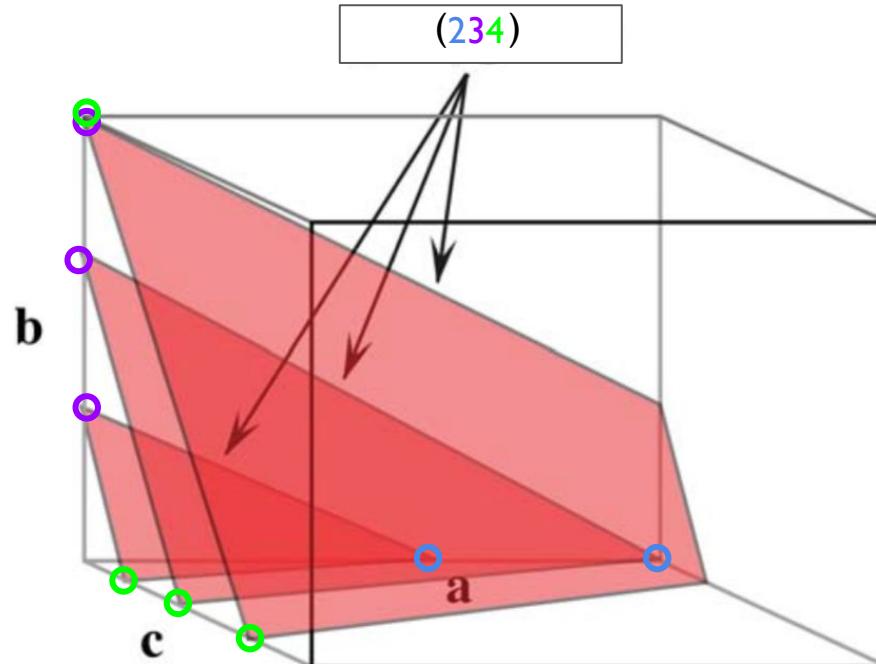


Source: Rhodes (2010)

X-ray crystallography

Miller indices (cont'd)

What will be the h and k indices of the group of planes in this example?

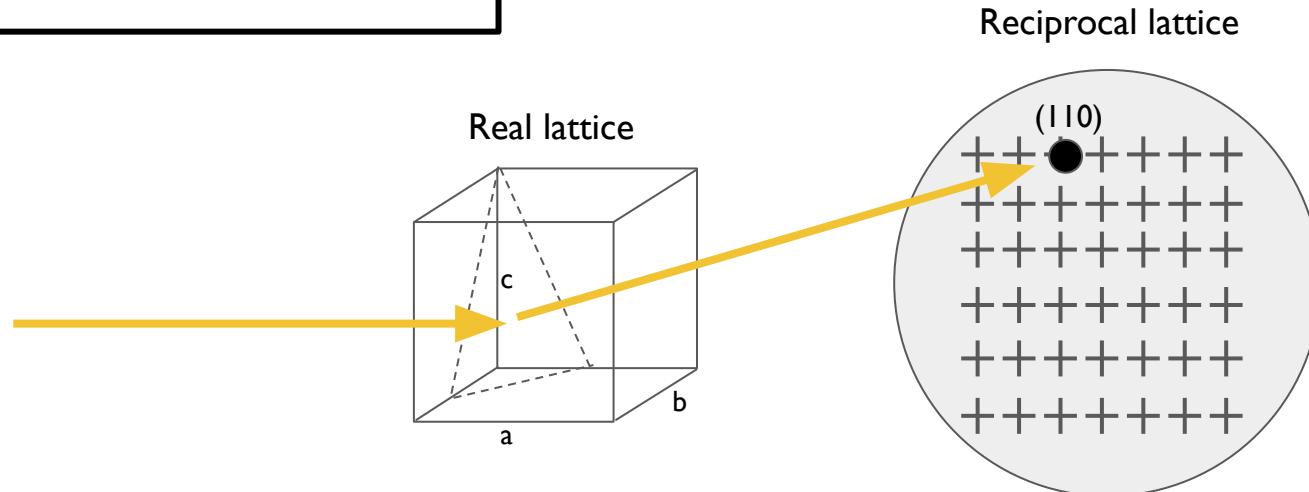


Source: Rhodes (2010)

X-ray crystallography

Significance of the Miller planes

Planes or plane groups produce **reflections** on the detector whose **intensities correspond to the electron density** on these planes.

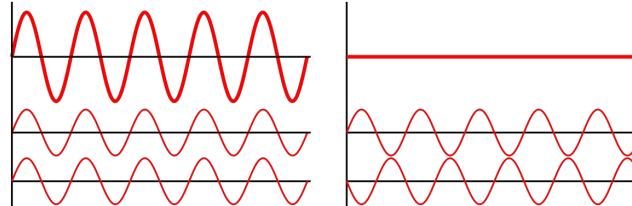


Why isn't the detector covered in infinitely many reflections?

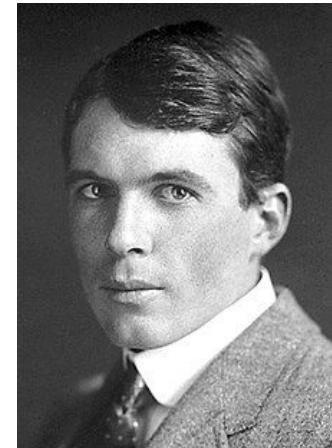
X-ray crystallography

Bragg's law (1913)

- Light behaves as a **wave** (Young's double-slit experiment of 1801)
 - **Constructive** and **destructive** interference



Source: Wave interference, Wikipedia



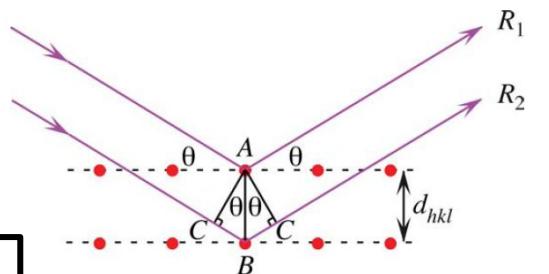
Lawrence Bragg
(1870-1971)

- Bragg's model of diffraction states that a reflection is only produced by a plane if

$$2d \sin \theta = n\lambda$$

plane spacing (resolution) angle of incidence wavelength of X-ray

All rays which dissatisfy this equality will cancel out via destructive interference!

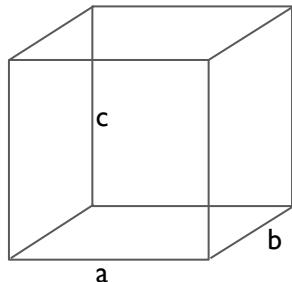


Source: Rhodes (2010)

X-ray crystallography

Relationship between the pattern and the structure

Scattering object



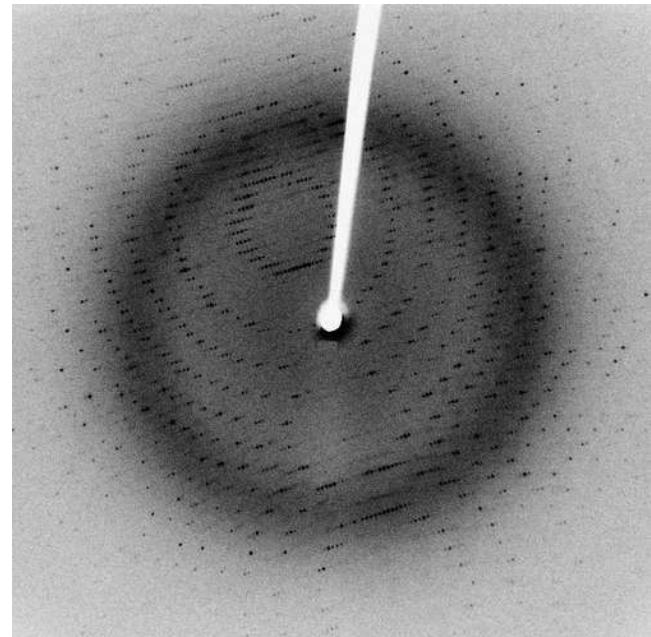
Fourier transform

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

Inverse Fourier transform

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx + ky + lz)}$$

Diffraction pattern



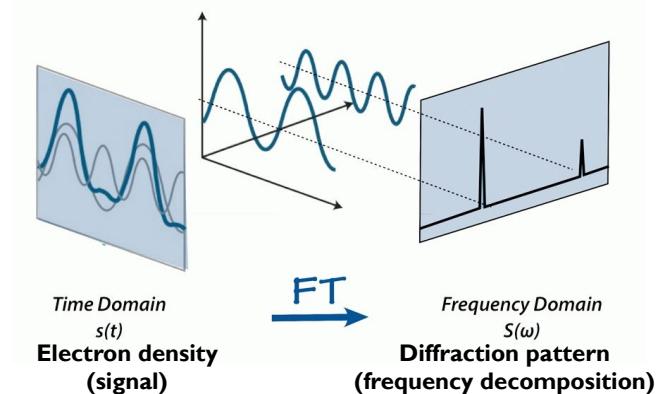
Source: <https://www.britannica.com/science/X-ray-diffraction>

Two different perspectives/views of same thing!

X-ray crystallography

Relationship between the pattern and the structure

- **Fourier transform**
 - Mathematical operation (transform)
 - Maps a signal from *time domain* to the *frequency domain*
 - Applications in physics, engineering, computer science, etc.
- Diffraction pattern is the **Fourier transform** of the 3D protein
- Electron density is the **inverse Fourier transform** of the diffraction pattern
- Every reflection in the diffraction pattern is a contribution (term in a sum) that comprises the entire electron density map



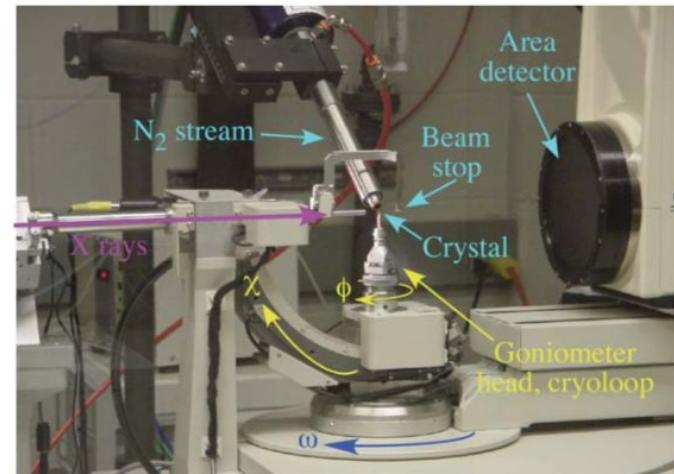
Source: <https://aavos.eu/glossary/fourier-transform/>

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

X-ray crystallography

Data collection process

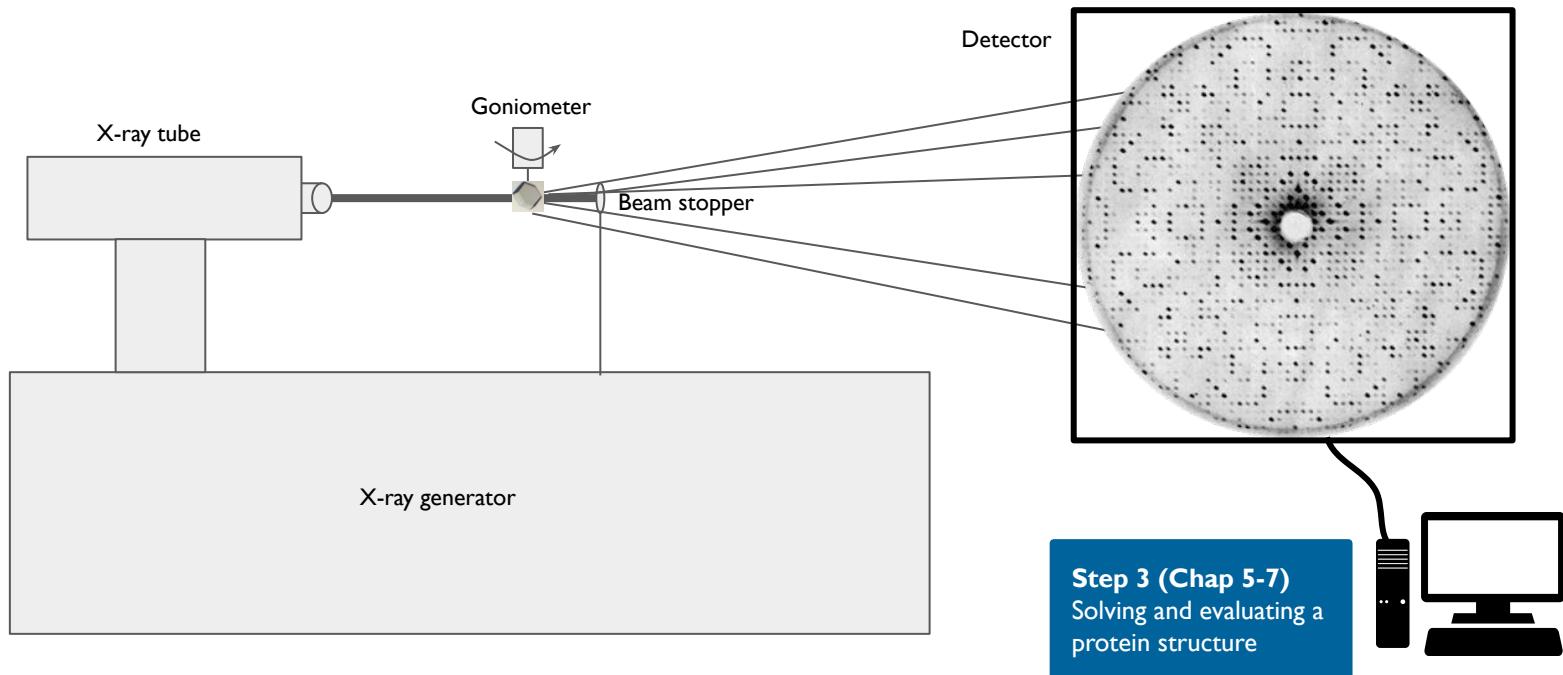
- Crystal is **screened**
 - Crystal diffract well?
 - Smallest d-spacing (maximum resolution)
- Decision collecting data **in-house** or at a **synchrotron**
 - Depends on resolution in-house
- Crystals **“fished” out** of solution and mounted on goniometer
- X-rays are shot onto the crystal
 - Angles between 0 and 180° collected (**why?**)
 - Crystal may need replacement
- **Result:** collection of images representing the entire reciprocal lattice of the unit cell



Source: Rhodes (2010)

X-ray crystallography

The principal steps of an X-ray crystallography experiment



X-ray crystallography

The phase problem

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i (hx + ky + lz - \alpha'_{hkl})}$$

Electron density at coordinates (x,y,z) in unit cell

Structure factor (amplitude) of hkl plane

hkl plane (frequency)

Phase

A diagram showing the components of the electron density equation. It consists of four parts: 1) 'Electron density at coordinates (x,y,z) in unit cell' with a curved arrow pointing to the term $\rho(x, y, z)$. 2) 'Structure factor (amplitude) of hkl plane' with a curved arrow pointing to the term $|F_{hkl}|$. 3) 'hkl plane (frequency)' with three arrows pointing to the term $e^{-2\pi i (hx + ky + lz)}$. 4) 'Phase' with a curved arrow pointing to the term $-\alpha'_{hkl}$. Below the first three components are green checkmarks, while the fourth component has a large red X.

X-ray crystallography

Solutions to the phase problem

Method 1: Molecular replacement phasing

- Leveraging information known about molecules that generated a similar diffraction pattern whose phase is known
- Use of homologous protein, protein-ligand complex, mutant, etc.

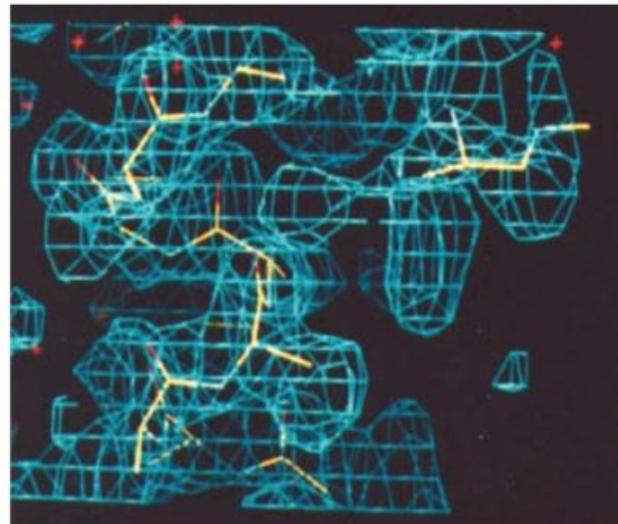
Method 2: Multiwavelength anomalous scattering (MAD) phasing

- Crystallization of the same protein in *E. coli* grown in media rich in selenomethionine
- Formation of identical crystals but for the selenomethionine
- Heavy atoms are sensitive to ray wavelength
- Use of multiple X-ray wavelengths
- Formation of reflections that are slightly different between wavelengths
- Difference between reflections holds phase information

X-ray crystallography

Model building

- Electron density map is obtained from the inverse Fourier transform
- Pockets that can be filled out with amino acids
- Computer-generated initial model
 - Further refined by crystallographer
- Adjustments
 - Amino acids
 - Geometry of bonds
 - Conformations
 - Dihedral angles
- Iterative changes to protein model



Source: Rhodes (2010)

X-ray crystallography

Optimization of an objective function

As we adjust the model, we attempt to **minimize an objective function Φ** which describes the **agreement between observed and calculated reflections**.

$$\Phi = \sum_{hkl} w_{hkl} (|F_o| - |F_c|)_{hkl}^2$$

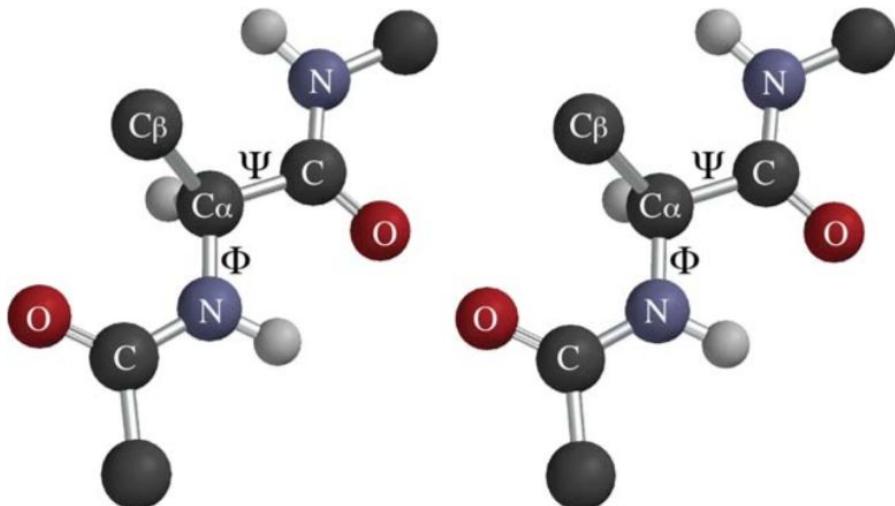
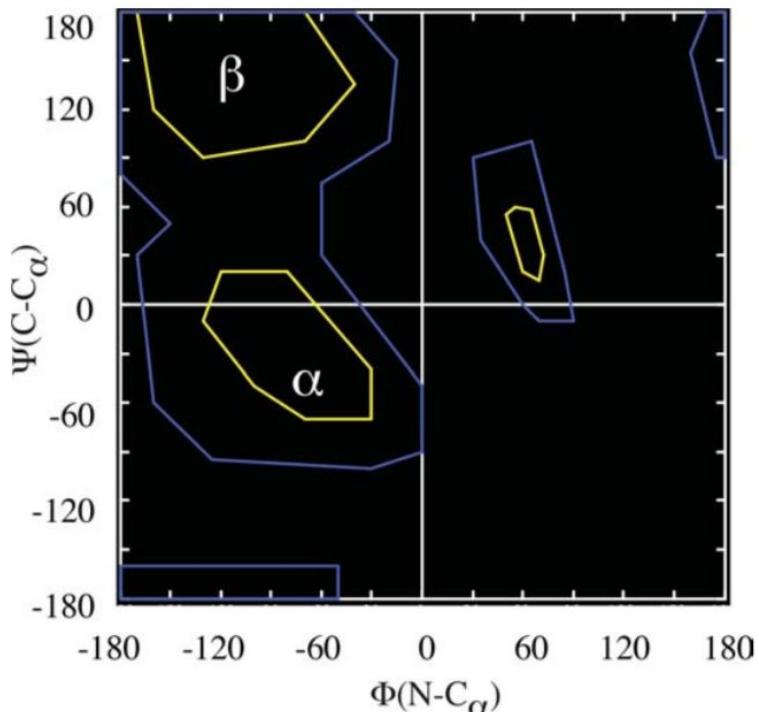
The diagram shows the objective function Φ as a sum over hkl indices. Three arrows point to different parts of the equation: one to w_{hkl} labeled "Weight of a reflection (based on reliability)", one to $|F_o|$ labeled "Intensity for plane hkl from diffraction pattern", and one to $|F_c|$ labeled "Intensity for plane hkl from model".

In practice, we monitor the residual index R :

$$R = \frac{\sum ||F_{obs}| - |F_{calc}|||}{\sum |F_{obs}|}$$

X-ray crystallography

Assessing the quality of a model - Ramachandran plot



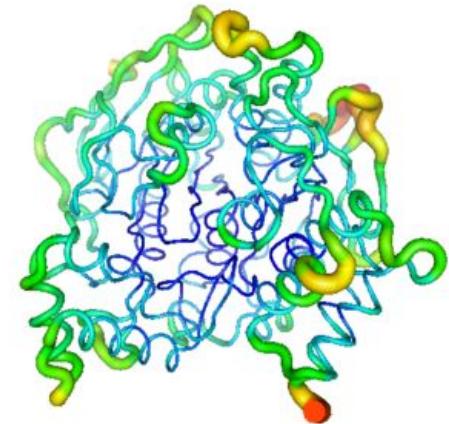
Source: Rhodes (2010)

X-ray crystallography

Assessing the quality of a model - B-factor

- Atom positions may vary slightly from a unit cell to the next
- Thermal motion: vibrational motion of atoms about their rest position
- Highly disordered regions associated with uncertainty of atom positions (may be dropped from model)
 - Protein terminals
 - Surface loop
 - Water molecules
- Measured with the B-factor (or thermal factor) in Å:

$$B_j = 8\pi^2 \{u_j^2\} = 79\{u_j^2\}$$



Source: Proteopedia

Mean square displacement of atom j
about rest position

X-ray crystallography

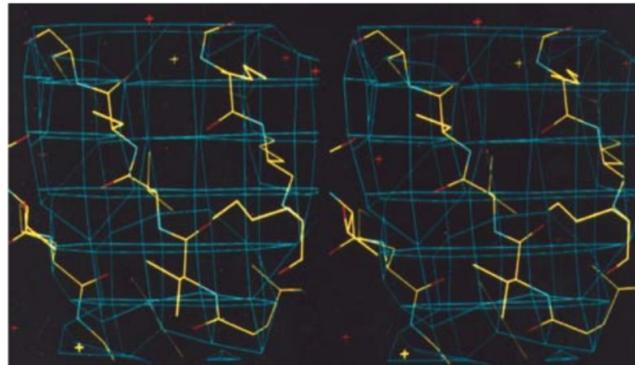
Assessing the quality of a model - Resolution

Resolution (Å)	Meaning
>4.0	Individual atomic coordinates meaningless. Secondary structure elements can be determined.
3.0 - 4.0	Fold possibly correct, but errors are very likely. Many sidechains placed with wrong rotamer.
2.5 - 3.0	Fold likely correct except that some surface loops might be mismodelled. Several long, thin sidechains (lys, glu, gln, etc.) and small sidechains (ser, val, thr, etc.) likely to have wrong rotamers.
2.0 - 2.5	As 2.5 - 3.0, but number of sidechains in wrong rotamer is considerably less. Many small errors can normally be detected. Fold normally correct and number of errors in surface loops is small. Water molecules and small ligands become visible.
1.5 - 2.0	Few residues have wrong rotamer. Many small errors can normally be detected. Folds are rarely incorrect, even in surface loops.
0.5 - 1.5	In general, structures have almost no errors at this resolution. Individual atoms in a structure can be resolved. Rotamer libraries and geometry studies are made from these structures.

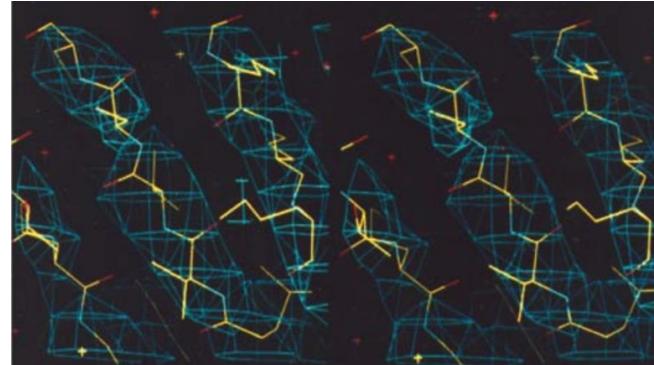
Source: [https://en.wikipedia.org/wiki/Resolution_\(electron_density\)#X-ray_crystallography](https://en.wikipedia.org/wiki/Resolution_(electron_density)#X-ray_crystallography)

X-ray crystallography

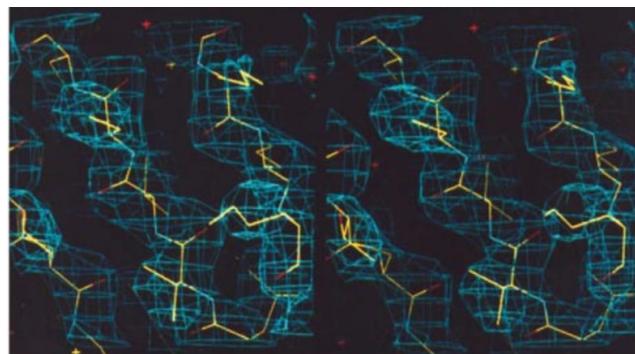
Assessing the quality of a model - Resolution



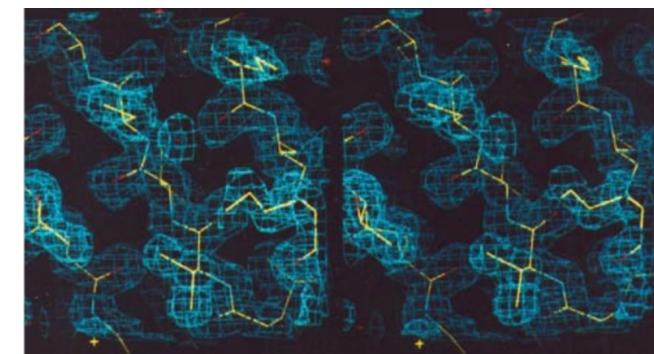
Resolution: 6 \AA



Resolution: 4.5 \AA



Resolution: 3 \AA



Resolution: 1.6 \AA

Source: Rhodes (2010)

X-ray crystallography

Reporting model statistics

Data collection	CjFurZn
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
a, b, c (Å)	35.74, 84.36, 123.63
Resolution	69.68–1.81
R _{sym}	6.3 (48.3) [†]
I/σI	5.3 (2.9)
Completeness (%)	100 (100)
Redundancy	3.4 (2.1)
Refinement	
Resolution	34.65–1.81
Reflections	34641
R _{work} /R _{free}	18.6/23.1
No. atoms	
Protein	2369
Zn ²⁺	2
Water	398
B-factors (Å ²)	
Protein	25.8
Zn ²⁺	13.7
Water	37.6
R.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.33
Molprobity score	1.59
Ramachandran favored (%)*	98.9

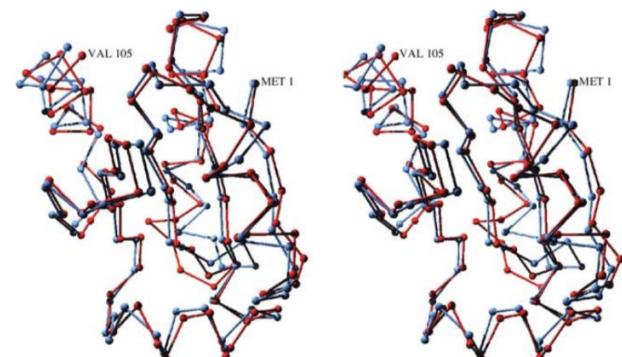
X-ray crystallography

How reliable is a crystal structure in general?

Good evidence that crystal structures are reliable representation of actual protein structure:

1. Preservation of enzyme activity in crystal!
2. Excellent agreement with NMR structures!
3. Structure identical for crystals obtained in different conditions of the same protein!
 - a. Crystal packing has minimal influence on structure

NMR vs. X-ray



Source: Rhodes (2010)

X-ray crystallography

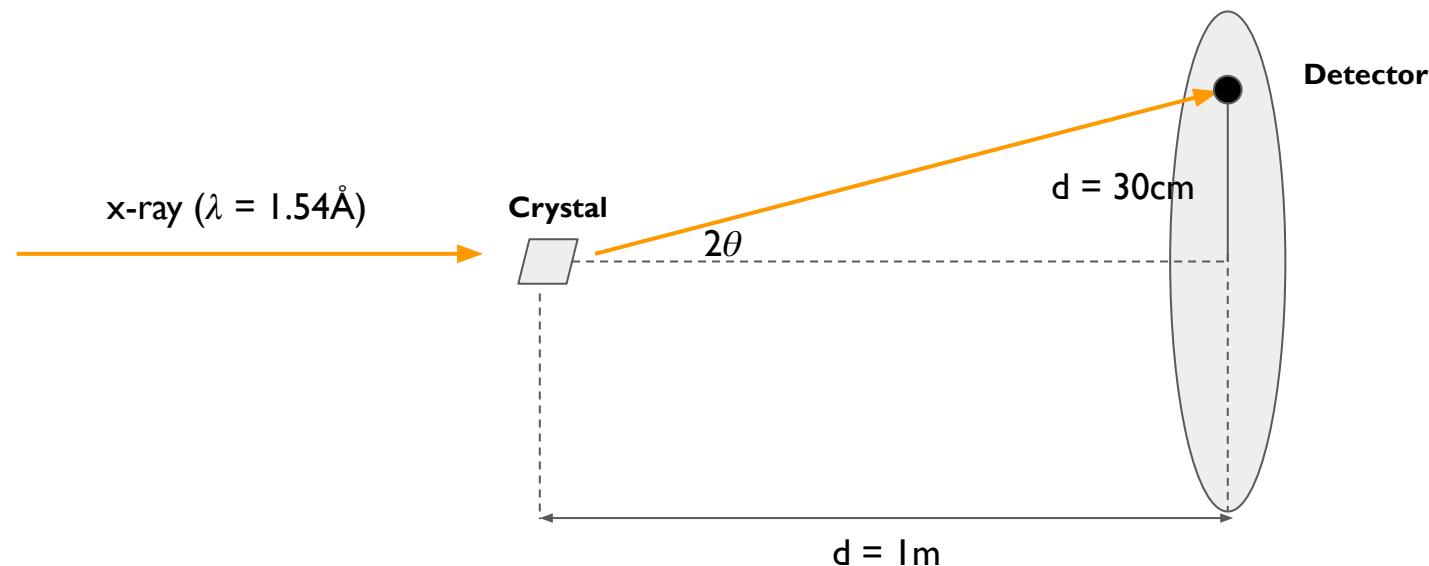
Take-home message

- **Getting protein crystals is a multistep process!**
 - Screening and optimization
 - Not all crystals diffract well
- **Reflexions on diffraction pattern correspond to electron density along planes in unit cell**
 - Inverse Fourier transform on pattern yields electron density map
- **Solving a structure is non-trivial**
 - Iterative process, fine adjustments required
 - Need to solve for phase
 - Minimize difference between observed and calculated reflections
- **Crystal structures are, in general, excellent representations of structure in solution**

Practice Question

X-ray crystallography - Practice question

I am running a screening crystallography experiment. I am using a typical x-ray beam with a wavelength of 1.54\AA . I notice that the farthest reflection from the center of the detector is 30cm. In my setup, the detector is 1m away from the goniometer (crystal mount). What is the maximal resolution that I can expect from this crystal? What will this kind of resolution tell me about my protein structure?



Practice Question

X-ray crystallography - Practice question (Solution)

If you look carefully at the slide on Bragg's law, you will see that the angle of incidence and reflection are the same, so the total angle of the reflection generated from a beam at a 0 deg angle is 2θ . We can solve for θ :

$$\theta = \frac{1}{2} \arctan(0.3m/1m) = 0.145 \text{ rad}$$

Using Bragg's law, we can determine the d , which corresponds to the space between the planes in the lattice, or the resolution! The closer the planes are together, the higher the resolution! We can rearrange Bragg's law and solve for d :

$$2d \sin(\theta) = n\lambda$$
$$d = n\lambda/(2 \sin(\theta))$$

There are, multiple possible values for n , but values larger than 1 will yield lower resolution reflections, and I am interested in the maximum achievable resolution, so:

$$d = \lambda/(2 \sin(\theta)) = (1.54 \text{ \AA})/(2 \sin(0.145 \text{ rad})) = 5.3 \text{ \AA}$$

This is a modest resolution. This allows for the determination of secondary structures and the overall conformation of the protein. The angles of the sidechain bonds cannot be determined with certainty.

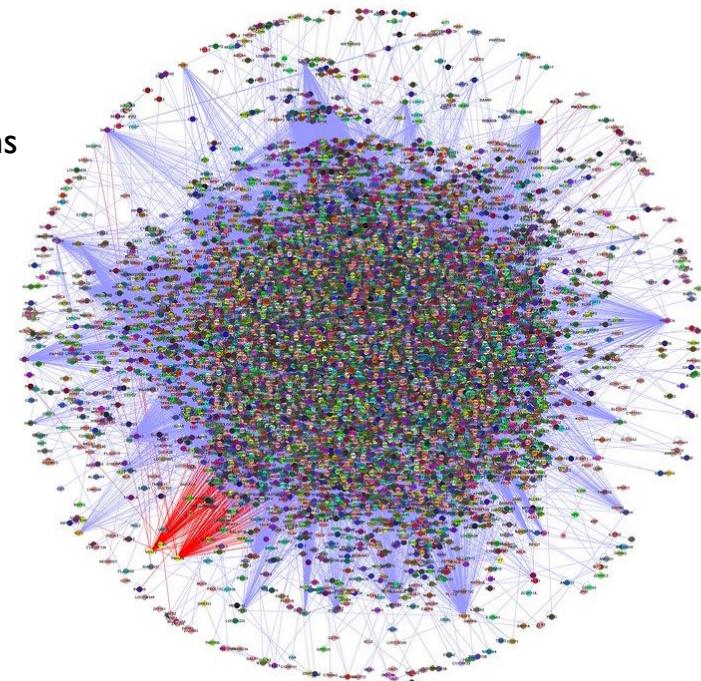
Part II - Computational biochemistry

Computational prediction of protein-protein interactions (PPIs)

Computational biochemistry

Protein-protein interactions (PPIs)

- **Why are protein-protein interactions important?**
 - Fundamental modality in cell pathways with enzymatic reactions
 - Aberrant interactions in disease due to mutations/translocations/etc.
- **How is this knowledge actionable?**
 - Disruption of PPIs with inhibitors can restore “normal” phenotype
 - Binding of inhibitor to binding interface
 - Binding in allosteric region to lock in a specific conformation that prevents interaction
- **Why is there a place for bioinformatics/computational approaches in this?**
 - $n_{\text{PPIs}} \approx 20,000(20,000 + 1)/2 = 200\text{M}$ possible interactions
 - Exhaustive identification of all interactions is intractable

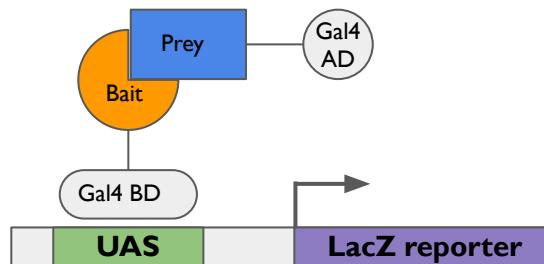


Visualization of the known interactome with Cytoscape
Source: <https://www.flickr.com/photos/andytrop/5234332602>

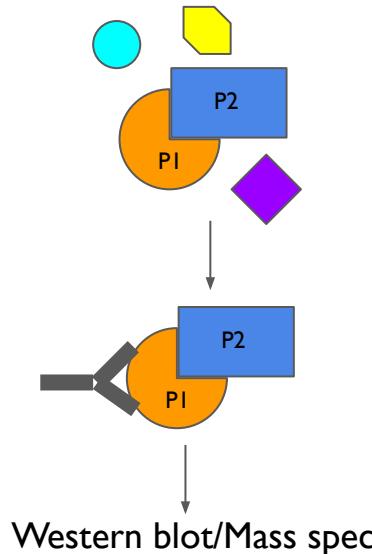
Computational biochemistry

Protein-protein interactions - Wet lab methods

There are multiple wet lab methods to validate protein-protein interactions. Which ones do you know?



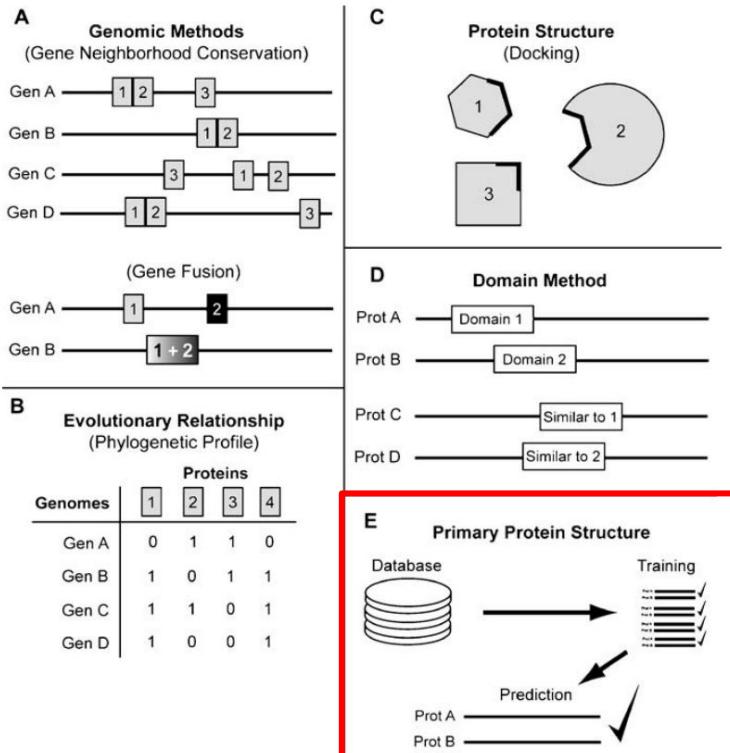
Yeast two-hybrid



Pull-down (e.g. co-IP)

Computational biochemistry

Protein-protein interactions - Families of prediction algorithms

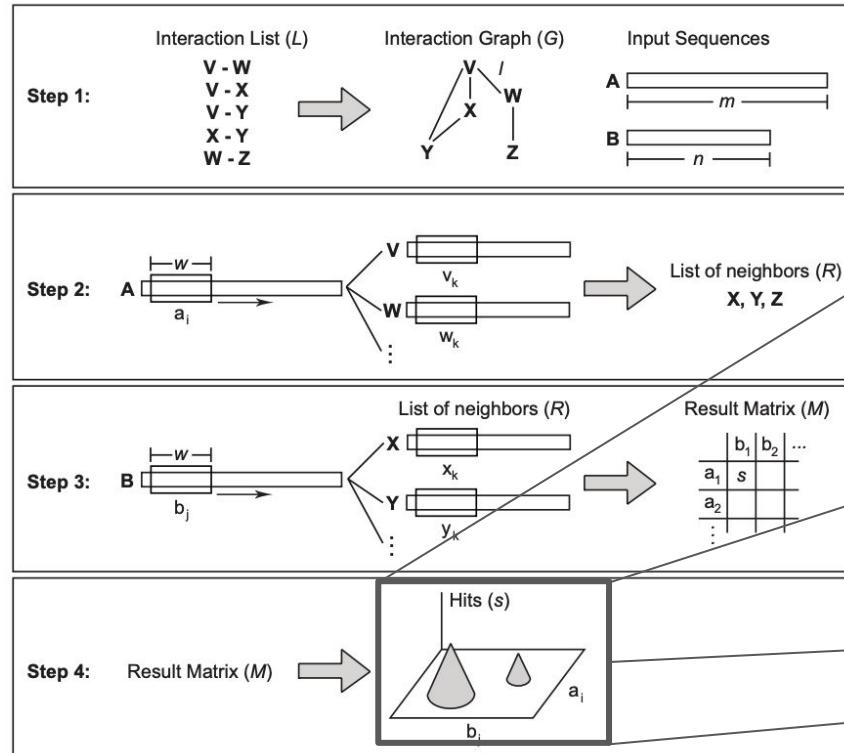


Source: Pitre et al. (2008)

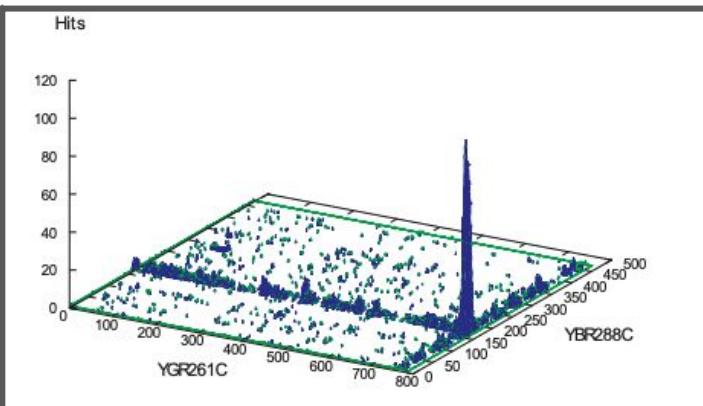
Fig. 1 The five categories of computation PPI methods: **A** Genes of proteins that are close in different genomes are predicted to interact. Proteins 1 and 2 are predicted to interact since the physical locations of their genes are in close proximity to each other in the genomes A, B and D. Two proteins are also predicted to interact if they combine (fuse) to form one protein in another organism. **B** Protein pairs with similar phylogenetic profiles in different genomes are predicted to interact. Proteins 1 and 4 are predicted to interact since they share the same phylogenetic profile. **C** Using the protein structures, docking methods will predict the best compatibility of their interacting regions. Proteins 1 and 2 are predicted to interact since they have the best fit. **D** If two proteins A and B known to interact share a pair of conserved domains and two other proteins C and D also share those same conserved domains, C and D are predicted to interact. **E** Using the primary protein structure and a database containing some other information (such as known interactions), it is possible to train an algorithm to predict protein–protein interactions

Computational biochemistry

Protein-protein interactions - Protein-protein Interactions Prediction Engine (PIPE)



Main idea: If two query proteins A and B contain regions that are similar to regions in proteins C and D in a database known to interact, there is evidence that A and B might interact.

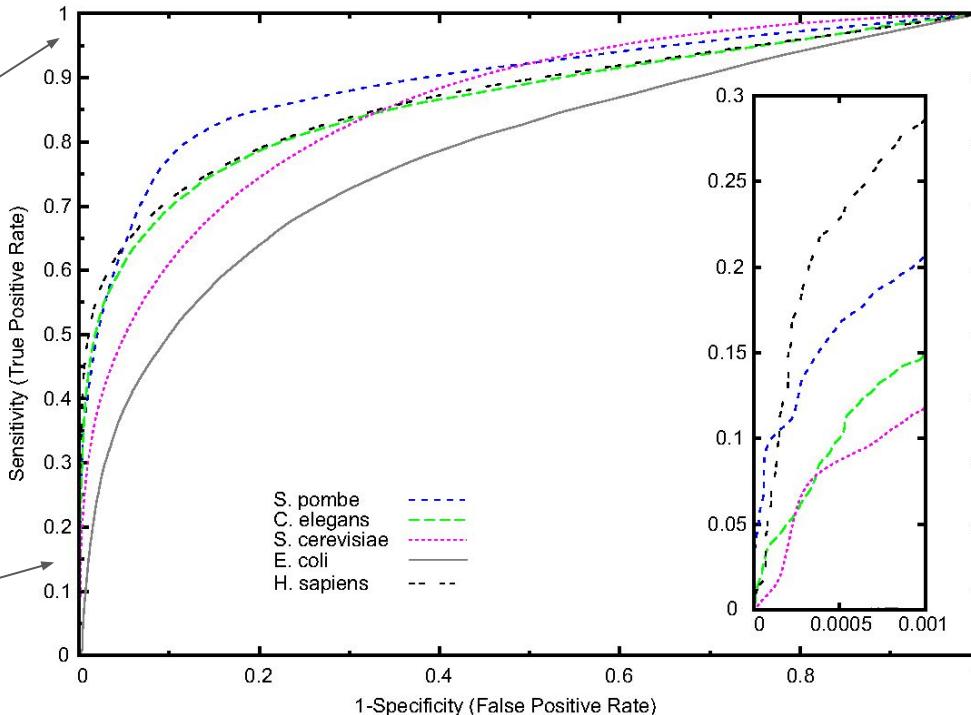


Source: Schoenrock et al. (2011)

Computational biochemistry

Protein-protein interactions - Performance (ROC curve)

Top left corner is perfect prediction



Interested in this region. Why?

Source: Pitre et al. (2006)

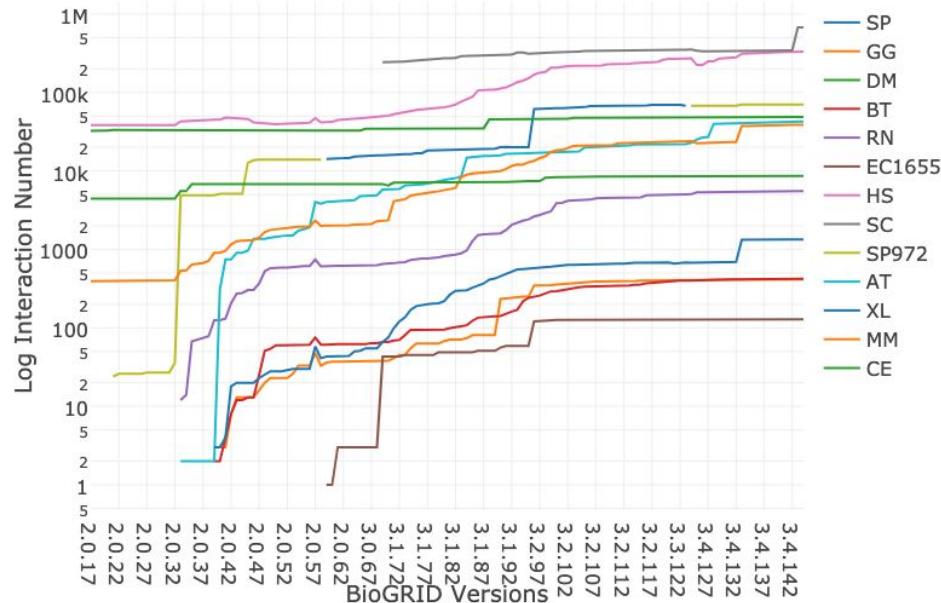
Computational biochemistry

Protein-protein interactions - POSITOME

- Lots of protein interaction data out there
 - Steadily growing
- Is it all high quality?
 - No, lots of false positives!
- Predicting protein-protein interactions with algorithms learning on poor quality data is a bad idea
 - Garbage in, garbage out
- What if we only selected high quality data?

POSITOME (Dick and Green, 2017)

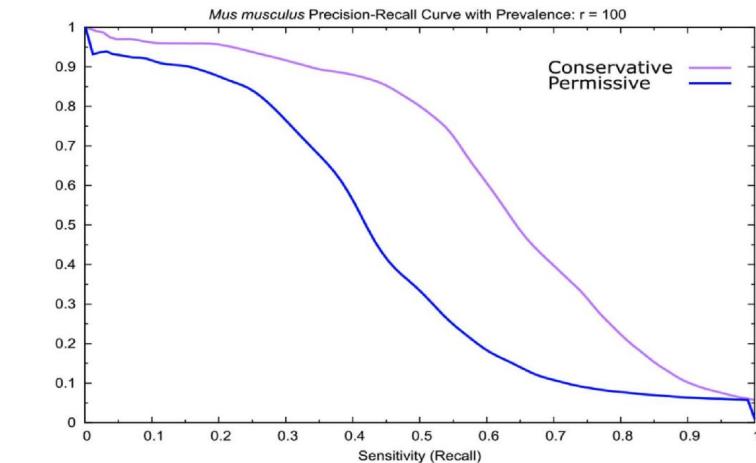
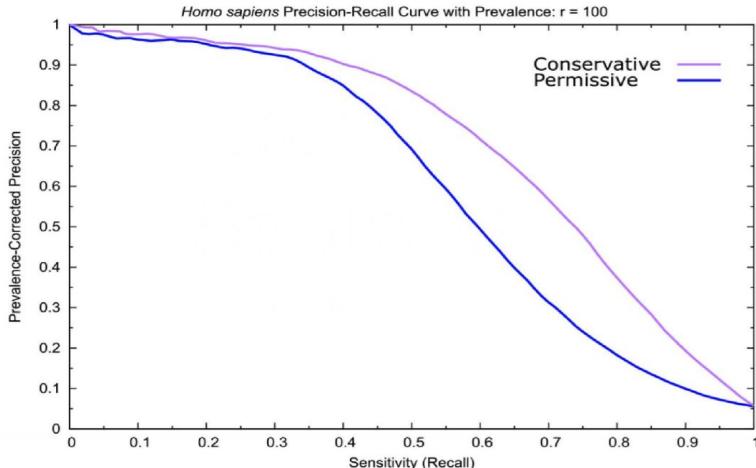
Historical BioGRID Version Growth



Computational biochemistry

Protein-protein interactions - POSITOME

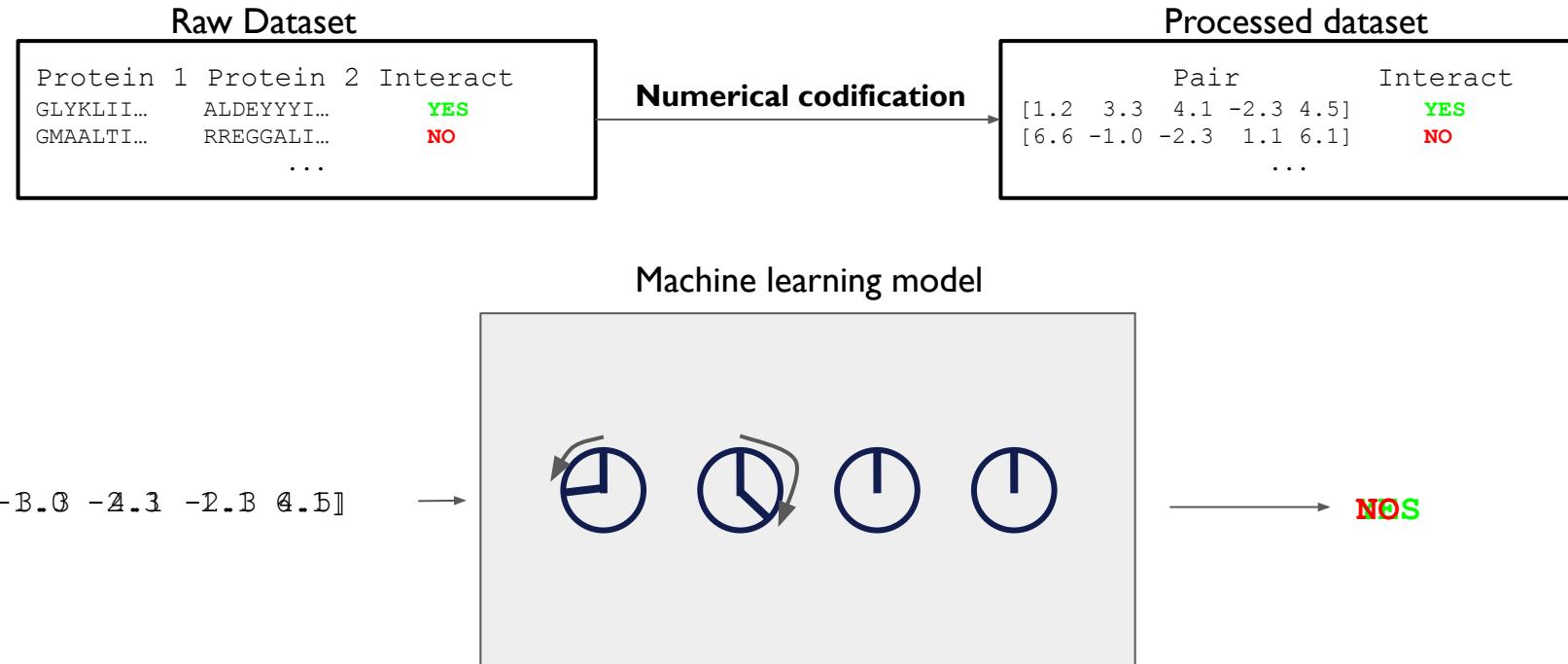
- Only use interactions characterized with specific, reliable methods (conservative sampling)
- Only use interactions with multiple lines of evidence
- Less data, but still, more accurate predictions



Source: Dick and Green (2017)

Computational biochemistry

Protein-protein interactions - Predicting with machine learning methods



Part II - Computational biochemistry

***In silico* drug design**

Computational biochemistry

Genetic algorithms

Genetic algorithm

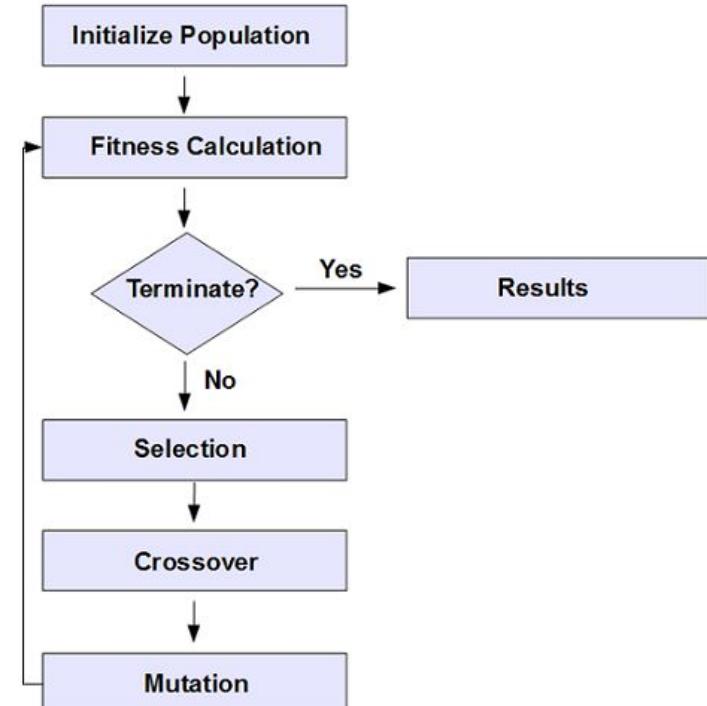
- Optimization algorithm that mimics evolution
- “Evolve” an optimal solution
- “Genetic-like” operations (*copy, crossover and mutate*)

Solution

- A solution can be a peptide, a protein structure, a DNA sequence, an algorithm, a route in Google Maps, etc.
- Must be able to define genetic operations over solutions

Fitness

- How “good” a solution is
- Arbitrarily defined by user
- Optimization criterion to maximize



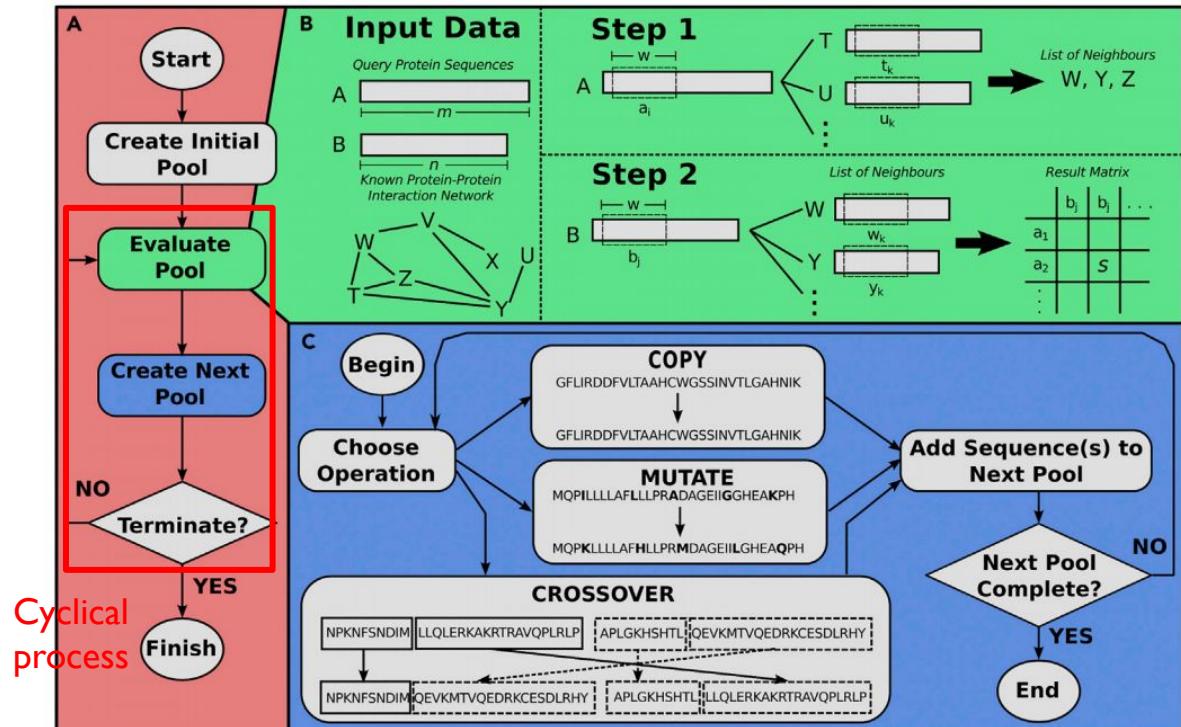
Source: <https://apacheignite.readme.io/docs/genetic-algorithms>

Computational biochemistry

Peptide inhibitor design - Overview of *In silico* Peptide Synthesizer (InSiPS)

- Designed by Andrew Schoenrock as a wrapper around PIPE
- **Idea:** use genetic algorithms to “evolve” peptide inhibitors for a target protein

What is a fit peptide?



Source: Burnside et al. (2019)

Computational biochemistry

Peptide inhibitor design - InSiPS (fitness)

$$f(p, t, \mathcal{N}) = [1 - \max(score(p, \mathcal{N}))] \times score(p, t)$$

The diagram illustrates the components of the fitness function. At the center is the equation $f(p, t, \mathcal{N}) = [1 - \max(score(p, \mathcal{N}))] \times score(p, t)$. Four arrows point from labels to different parts of the equation: one arrow from 'Set of all non-target proteins' points to the $\max(score(p, \mathcal{N}))$ term; another from 'Peptide' points to the variable p ; a third from 'Target protein' points to the variable t ; and a fourth from 'Scoring function (PIPE)' points to the entire equation.

We want to maximize both factors in the equation:

- 1) Opposite of maximum off-target score
- 2) On-target score

Computational biochemistry

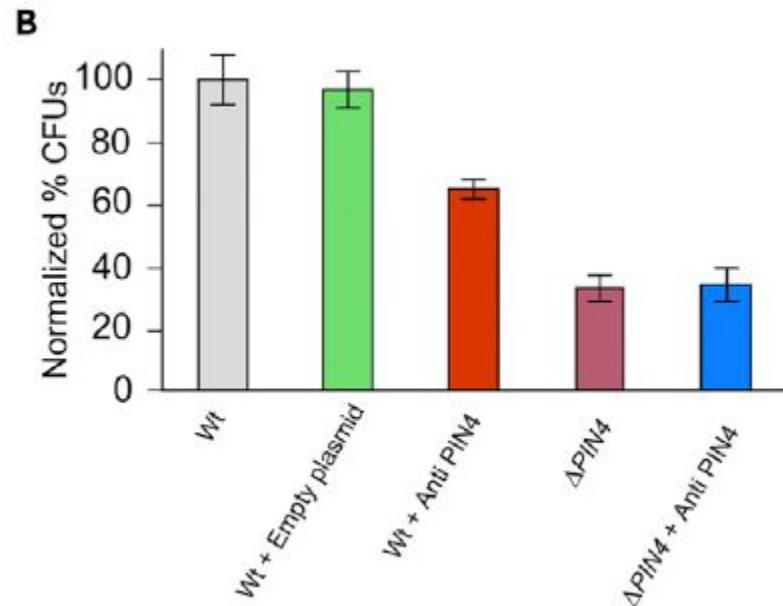
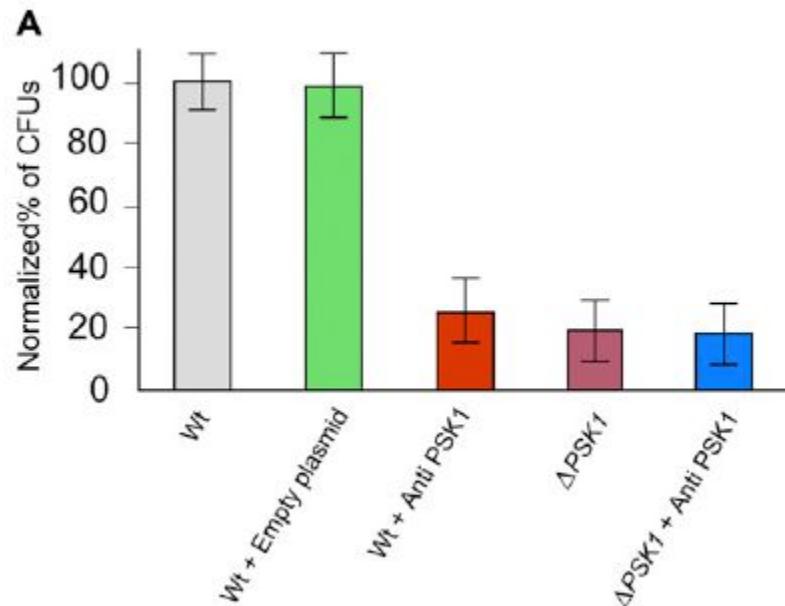
Peptide inhibitor design - InSiPS (results)

Synthetic Binding Protein	Fitness	Target Score	Max Non-Target Score	Max Non-Target	Average Non-Target Score	Closest Yeast Homolog to SBP
Anti-Psk1	0.465	0.718	0.352	Ubi4p	0.072	Mmp1p
Anti-Pin4	0.380	0.630	0.398	Cdc39p	0.0797	Esl1p
Anti-Rmd1	0.344	0.563	0.389	Sec14p	0.132	YAP1801p

Source: Burnside et al. (2019)

Computational biochemistry

Peptide inhibitor design - InSiPS (results)



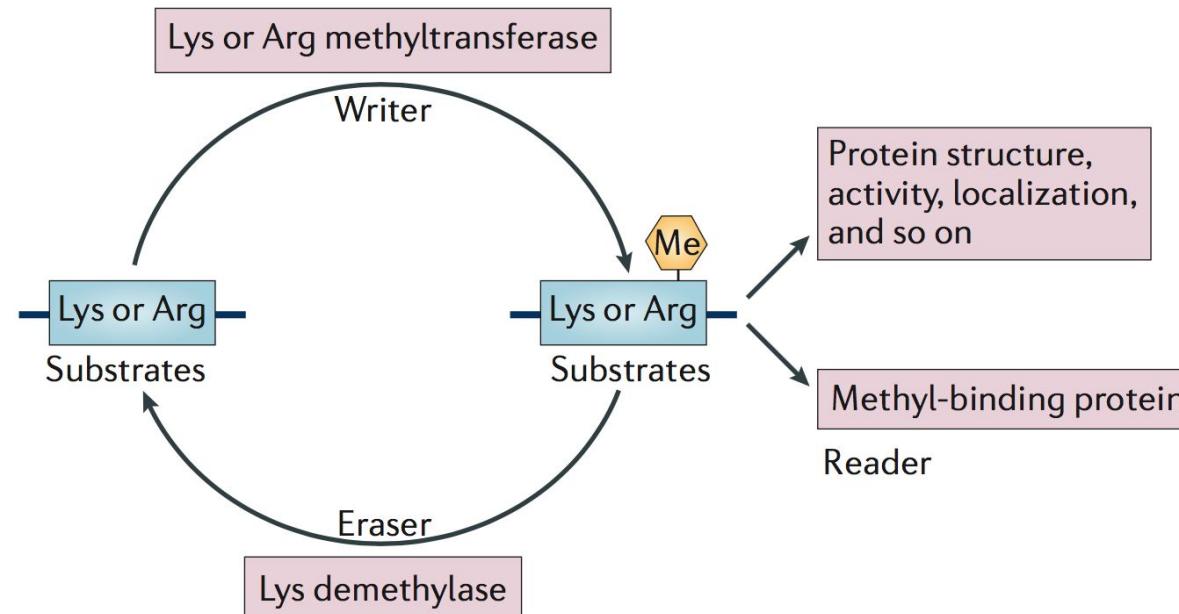
Source: Burnside et al. (2019)

Part II - Computational biochemistry

Prediction of lysine methylation sites

Computational biochemistry

Prediction of lysine methylation sites - Lysine methylation

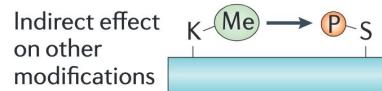
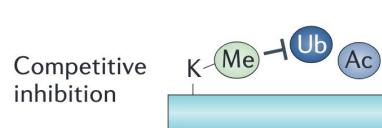


Source: Biggar and Li (2015)

Computational biochemistry

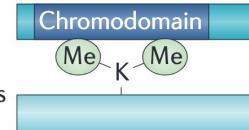
Prediction of lysine methylation sites - Role of methylation

a Other protein modifications

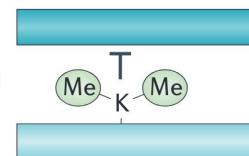


b Protein–protein interactions

Methylated-lysine-specific binding proteins

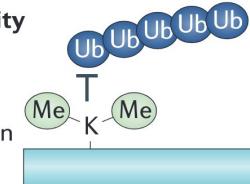


Inhibition of protein–protein interactions

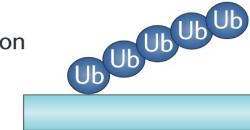


c Protein stability

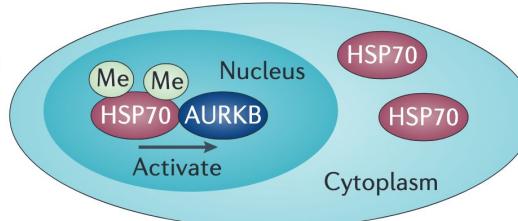
Inhibition of polyubiquitylation



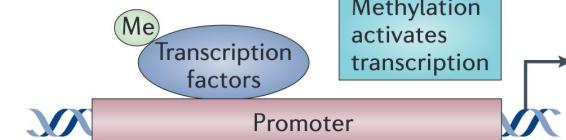
Protein degradation by proteasomes



d Subcellular localization



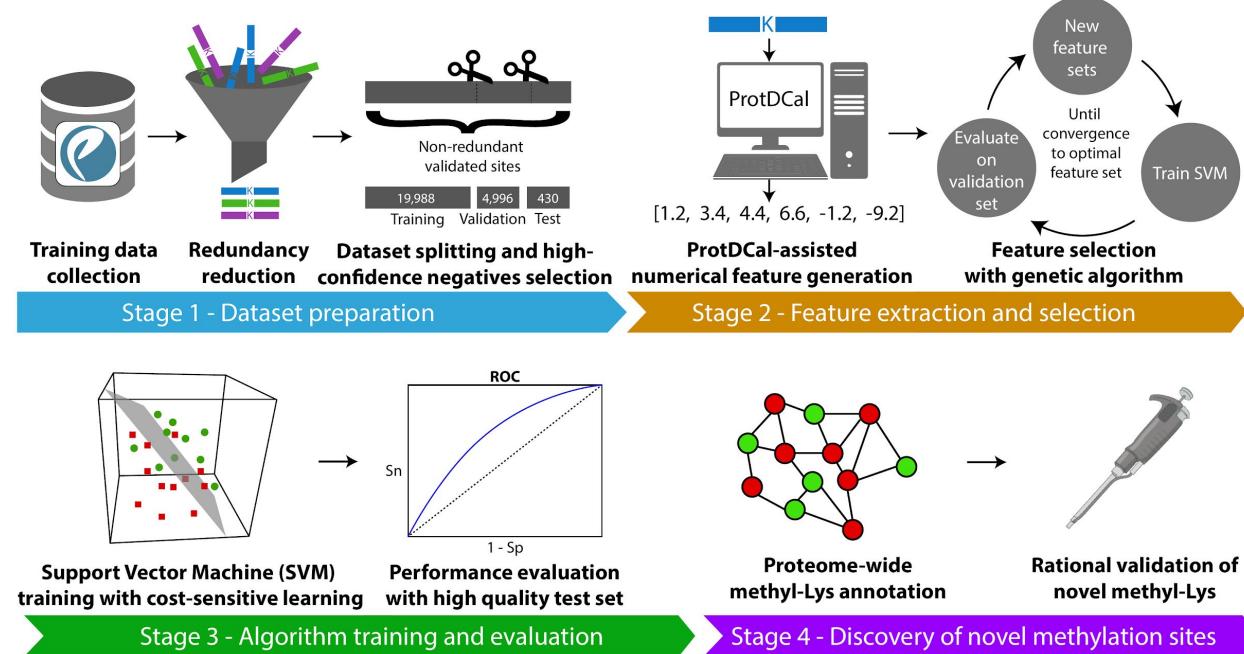
e Promoter binding



Source: Han et al. (2019)

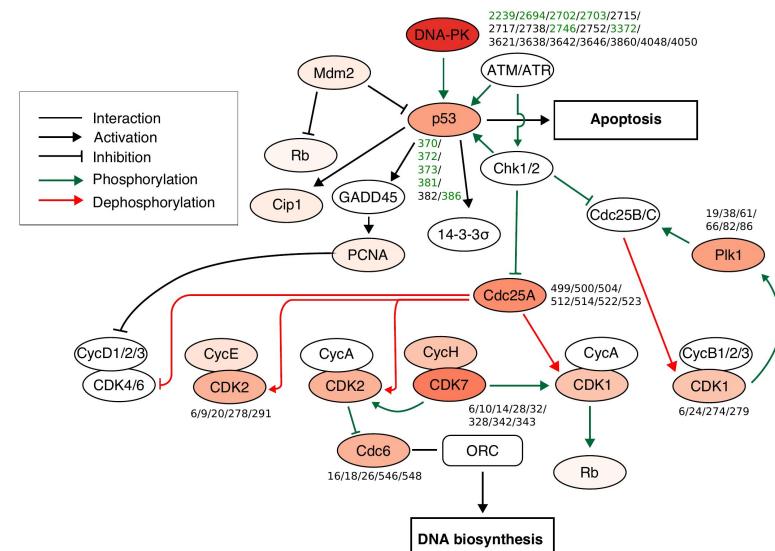
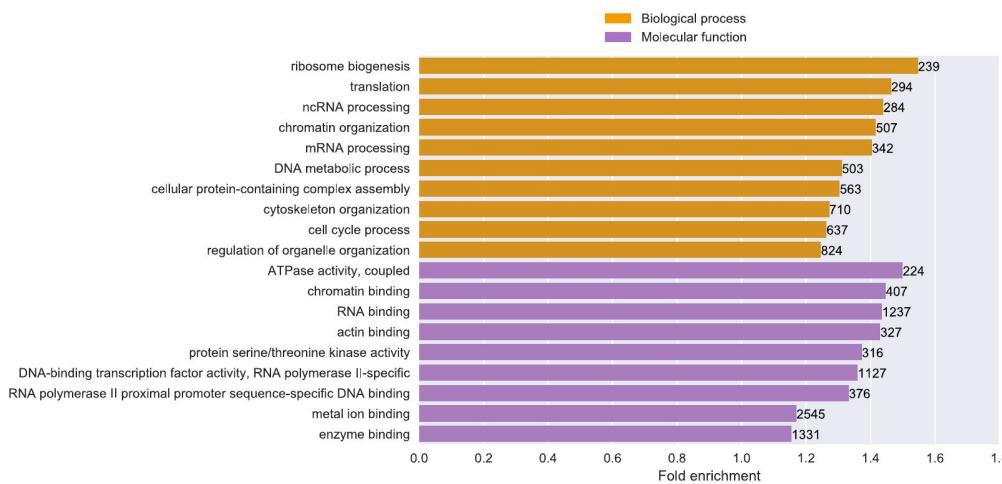
Computational biochemistry

Prediction of lysine methylation sites - MethylSight



Computational biochemistry

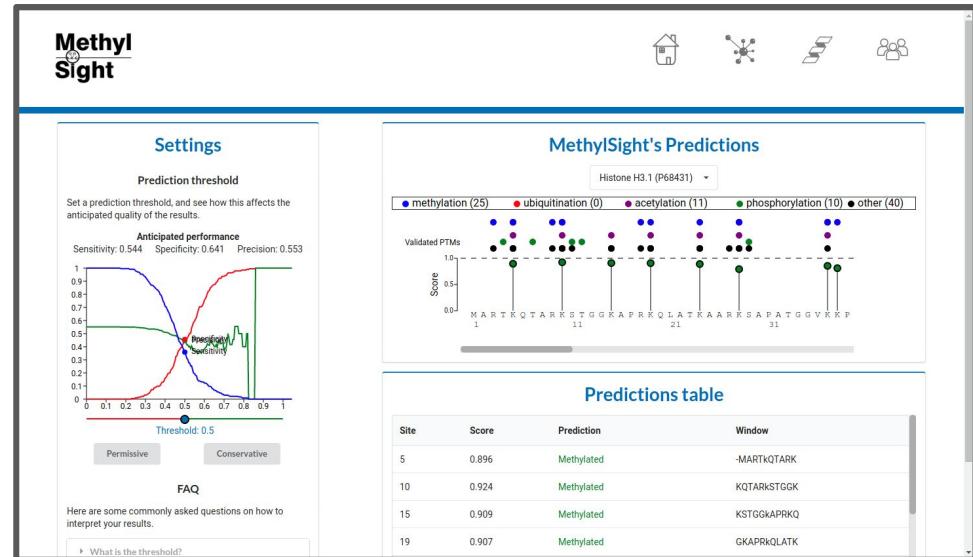
Prediction of lysine methylation sites - MethylSight (Results)



Computational biochemistry

Prediction of lysine methylation sites - Web server

- Computational methods useless used by biologists
- Many people choose to deploy their computational tools as **web servers**
 - Access through the software via a web interface
 - Interactive visualization of output



Computational biochemistry

Take-home message

- **Supports/guide laboratory experiments**
 - Cheaper and faster
 - Run on -ome scales (proteome, genome, metabolome)
- **Accelerates rate of discovery**
 - Rely on known PPIs (PTMs) to discover new ones
- **Machine learning is changing how we do computational biology**
 - Learn from existing data
 - Fast and streamlined (existing software)

References

- [1] Biggar, K.K., and Li, S.S.-C. (2015). Non-histone protein methylation as a regulator of cellular signalling and function. *Nature Reviews Molecular Cell Biology* *16*, 5–17.
- [2] Burnside, D., Schoenrock, A., Moteshareie, H., Hooshyar, M., Basra, P., Hajikarimlou, M., Dick, K., Barnes, B., Kazmirschuk, T., Jessulat, M., et al. (2019). In Silico Engineering of Synthetic Binding Proteins from Random Amino Acid Sequences. *IScience* *11*, 375–387.
- [3] Dick, K., Dehne, F., Golshani, A., and Green, J.R. (2017). Positome: A method for improving protein-protein interaction quality and prediction accuracy. In 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), (Manchester, United Kingdom: IEEE), pp. 1–8.
- [4] Han, D., Huang, M., Wang, T., Li, Z., Chen, Y., Liu, C., Lei, Z., and Chu, X. (2019). Lysine methylation of transcription factors in cancer. *Cell Death Dis* *10*, 290.
- [5] Pitre, S., Dehne, F., Chan, A., Cheetham, J., Duong, A., Emili, A., Gebbia, M., Greenblatt, J., Jessulat, M., Krogan, N., et al. (2006). PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. *BMC Bioinformatics* *7*, 365.
- [6] Pitre, S., Alamgir, M., Green, J.R., Dumontier, M., Dehne, F., and Golshani, A. (2008). Computational methods for predicting protein-protein interactions. *Adv Biochem Eng Biotechnol* *110*, 247–267.
- [7] Schoenrock, A., Dehne, F., Green, J.R., Golshani, A., and Pitre, S. (2011). MP-PIPE: a massively parallel protein-protein interaction prediction engine. In Proceedings of the International Conference on Supercomputing - ICS '11, (Tucson, Arizona, USA: ACM Press), p. 327.

Merci!
Thank you!

francoischarih@sce.carleton.ca
<http://charih.ca>