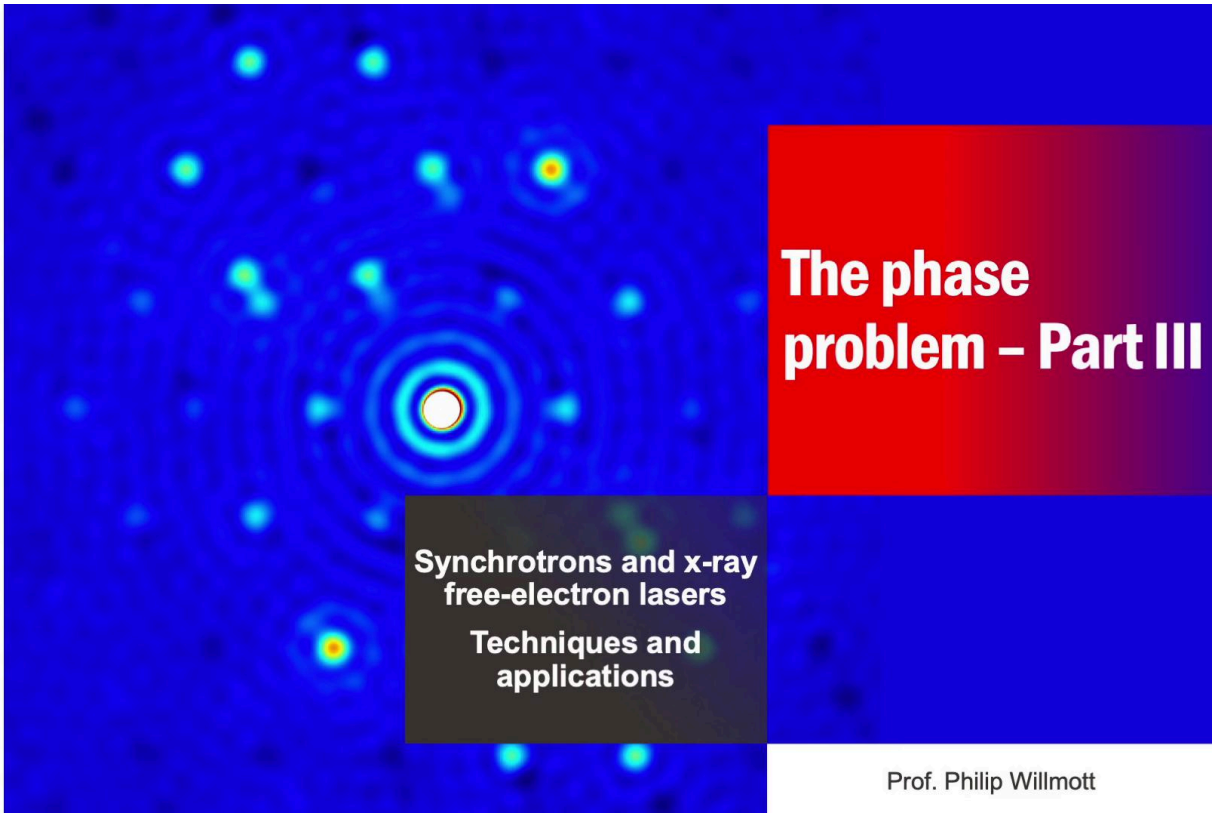
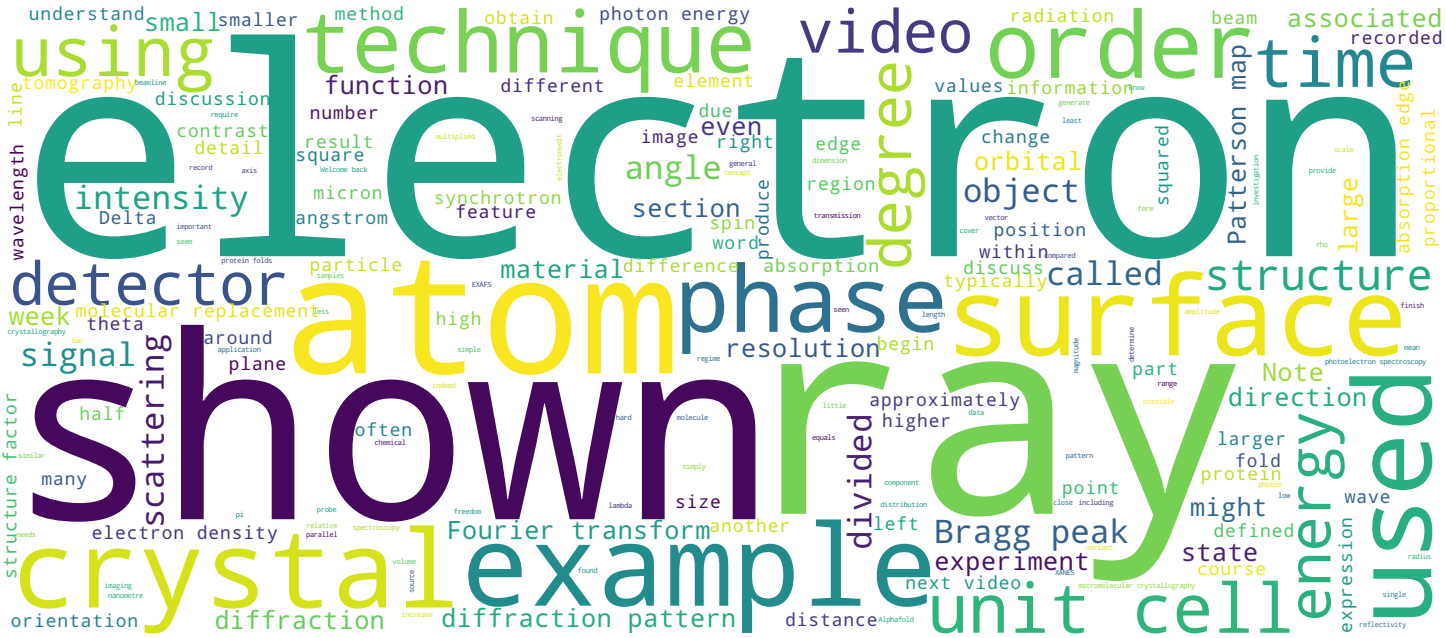


École polytechnique fédérale de Lausanne

Prof. Philip Willmott



Search MOOC

Video

EPFL

Contents and objectives of this video

- The Patterson map
- Molecular replacement

Hello again. We're nearing the end of our discussions on phasing in crystallography, especially macromolecular crystallography. The workhorse of phasing in MX is molecular replacement. In order to understand how we make our lives simpler when using this technique, we first need to understand the concept of the so-called Patterson map. These two aspects are the topics in this video.

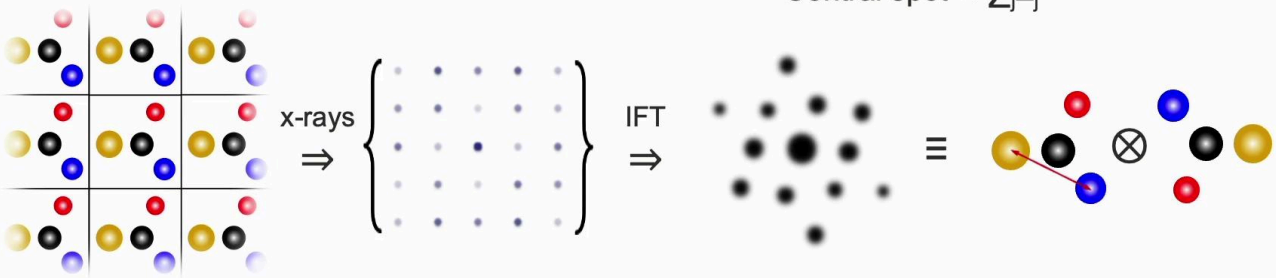
Notes

Summary

0m 05s

The Patterson map

- What information can we extract from the Bragg-peak intensities alone (i.e., without phases)?
- The inverse-FT of the Bragg-peak intensities yields the relative distances and angles between atoms in the unit cell
- Equivalent to convolution of electron density $\rho(r)$ with its inverse $\rho(-r)$
  - PMs are centrosymmetric
- For N atoms
  - $N(N-1) + 1$ maxima $\simeq N^2$
  - Atoms i and j: $I_{ij} \propto Z_i Z_j$
  - Central spot $\propto \sum_j Z_j^2$



See also derivation of the Patterson function in the supplementary text "The Patterson function"

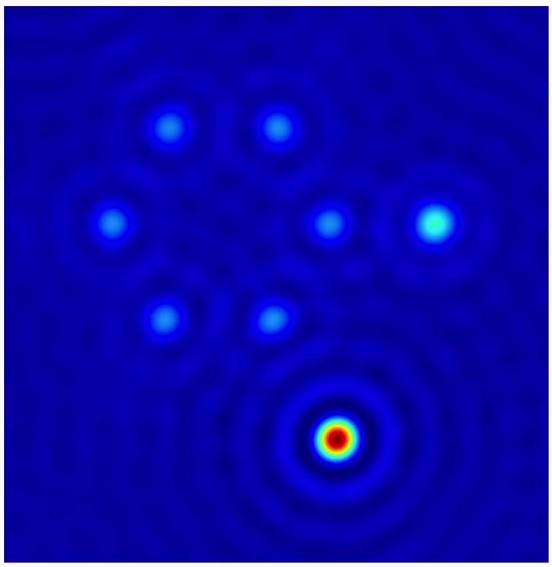
A priori, we don't know the phases of the structure factors in crystallography. I've said this many times before; this is our phase problem. But can we extract any structural information from the Bragg-peak intensities alone without the phase information? Not surprisingly, I wouldn't have posed this question if the answer were no. It turns out, and you can see the derivation in the supplementary text, "the Patterson function", that the inverse Fourier transform of the Bragg-peak intensities yields the relative distances and angles between the atoms within the unit cell, as shown here for the simple case of a unit cell containing only four atoms. This result is equivalent to the convolution of the electron density with its inverse. As a result, Patterson maps are centrosymmetric. For a unit cell containing N atoms, there are N times N minus one maxima plus a central maximum. For each pair of atoms i and j, the intensity of the associated peak is proportional to the product of their atomic numbers. For the central spot, its intensity is proportional to the sum of the square of each of the atom's Z values.

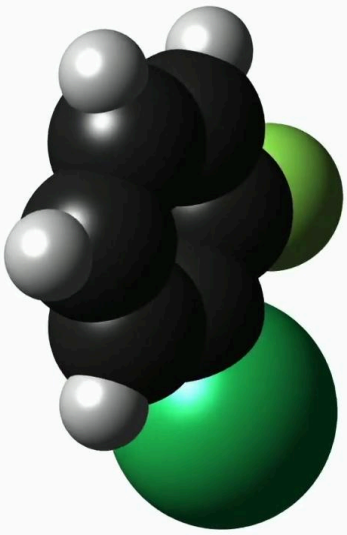
Notes

Summary

0m 36s

Time Out - The Patterson map





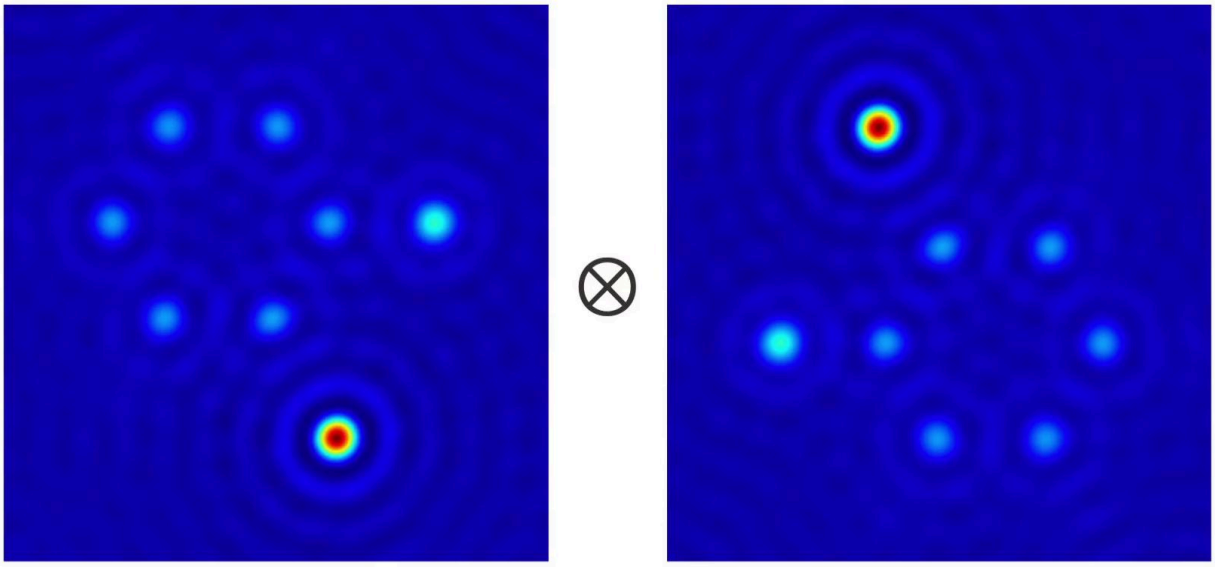
Let's consider an example based on the chlorofluoro benzene structure we met earlier this week already.

Notes

Summary

1m 59s

Time Out - The Patterson map



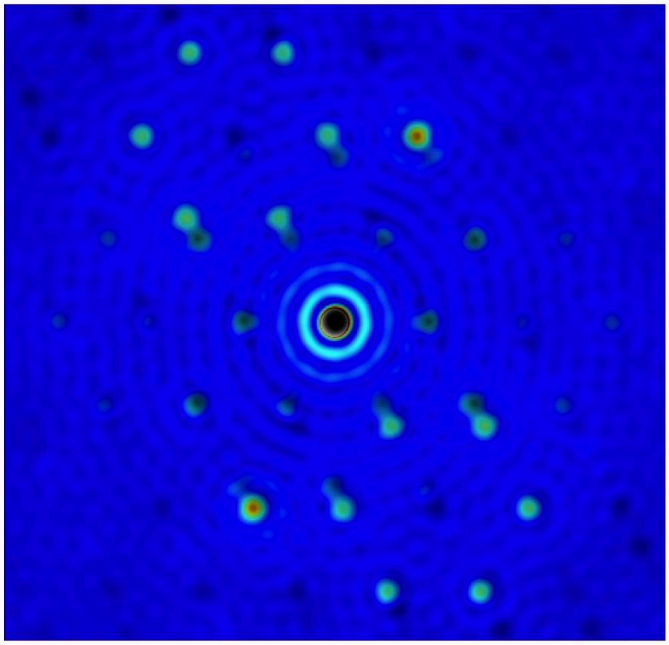
We convolute the electron density map with its inverse and obtain the pattern on the left.

Notes

Summary

2m 08s

The Patterson map



On the right is the Fourier transform of the set of intensities that were used to generate the original electron density map. We see that these correspond one to one. Note that the Patterson function of protein structures are extremely crowded affairs, consisting of normally many millions of peaks, making the whole thing appear very noisy. Any attempt to extract relative coordinates directly are doomed to failure.

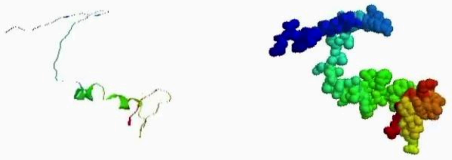
Notes

Summary

2m 18s

Molecular replacement

- Protein “folds”
  - How “secondary structures” (α-helices, β-sheets) arrange themselves in space
  - One fold can have different amino-acid sequences (“homologous”)
- Limited number of folds in nature
  - ca. 1400 discovered to date
  - Last new entry in 2008
- Use this information to glue together “blocks” of folds.
- What are their relative orientations and positions?



Folding of streptococcal protein G. Courtesy: Nancy Amato
https://parasol.tamu.edu/groups/amatogroup/research/folding/proteinG.php

Courtesy: Arto Alanenpää, Creative Commons

T. Hegedüs *et al.*, doi: 10.1007/s00018-021-04112-1

We now understand the Patterson map. How can we exploit this function in so-called molecular replacement? First, we need to understand protein folds in a little more detail. Protein folds describe how secondary structures such as α-helices and β-sheets arrange themselves in space. Note that a given fold can be generated by more than one amino acid sequence in a homologous manner. For once, nature plays on our side, insofar that it seems as if there are a relatively limited number of folds that she produces. To date, some 1,400 folds have been identified, the most recent discovery being in 2008. It may emerge that a whole new menagerie of folds are discovered in the future, but these would most likely be associated with membrane proteins, which to date, have been difficult to synthesise as crystals on account of their hydrophobic nature. It seems as if at least some new folds are being predicted by Alphafold 2. It remains to be seen how many more will be discovered. See the link below for a deeper discussion of this topic. Nonetheless, for those proteins that lend themselves to crystallography, it would appear that we have a near complete set of protein folds.

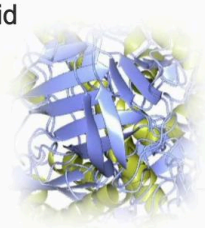
Notes

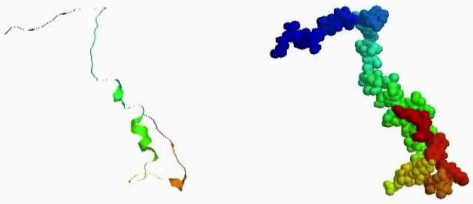
Summary

2m 51s

Molecular replacement

- Protein “folds”
  - How “secondary structures” (α-helices, β-sheets) arrange themselves in space
  - One fold can have different amino-acid sequences (“homologous”)
- Limited number of folds in nature
  - ca. 1400 discovered to date
  - Last new entry in 2008
- Use this information to glue together “blocks” of folds.
- What are their relative orientations and positions?





Folding of streptococcal protein G. Courtesy: Nancy Amato
https://parasol.tamu.edu/groups/amatogroup/research/folding/proteinG.php

Courtesy: Arto Alanenpää, Creative Commons

T. Hegedüs *et al.*, doi: 10.1007/s00018-021-04112-1

We can use these folds like they are modular elements, which we can glue together into the three-dimensional structure of the protein under investigation. To do this, we need to know their relative orientations and positions. This is where the Patterson map comes into play in so-called molecular replacement.

Notes

Summary

4m 21s

Molecular replacement

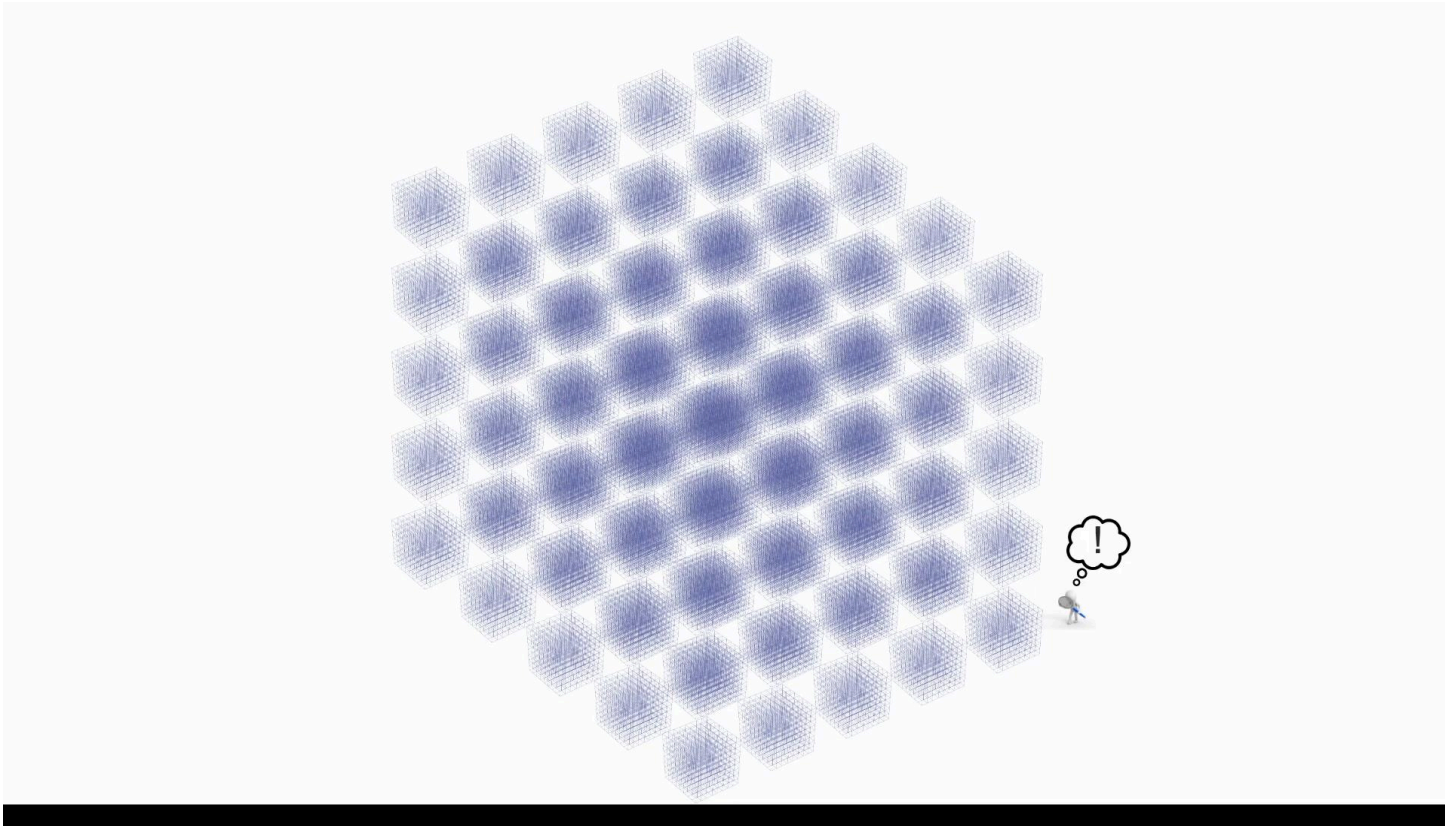
- Take coordinates of known fold
- Translate and rotate fold until simulated diffraction pattern most resembles the experimental data
- Six degrees of freedom (pitch, yaw, roll, x, y, z)
  - Without clever approach, this will take a long time! Very long!

Molecular replacement is a pragmatic model-building algorithm which works because there is now a large database of macromolecular structures from which reliable structural information can be drawn. One takes the internal coordinates of a known fold, even one that might be homologous, and places it within the known unit cell volume of the protein under investigation. One generates the diffraction pattern of this. This can be compared to the experimental diffraction pattern. There should be a high correlation between the two if the orientation and position of the fold are correct. However, there is essentially no chance that the initial orientation and position with the unit cell is correct. One, therefore, needs to check all six degrees of freedom of x, y, and z, and pitch, yaw, and roll. Without using a clever approach, probing this six-dimensional space would take a long time, extremely long.

Notes

Summary

4m 46s



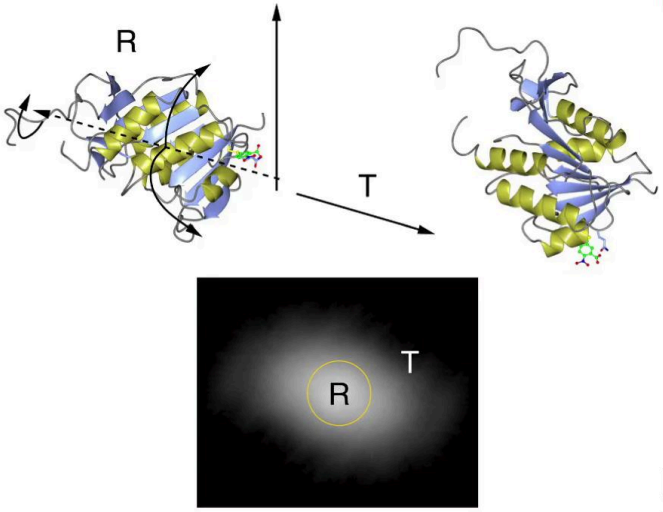
Just checking one parameter, that's relatively simple, while the time needed to check two degrees of freedom increases in a quadratic fashion. Things become exponentially worse as each degree of freedom is added, until the entire process becomes quite overwhelming. So for example, an operation that might take 0.1 seconds to probe 200 points in one dimension would require 0.1 times 200 to the 5 which is equal to 3.2 times 10 to the 10 seconds or 10 years to sample with equal resolution in all six dimensions.

Notes

Summary

5m 56s

Molecular replacement



- "Divide and conquer" using the Patterson map
  - Angular orientation (R) + position within unit cell (T)
  - First determine R
    - Interatomic spacings over distances ≤ size of fold
    - Compare and optimize (via three angle variables) central part of Patterson map with that of known fold
  - Then determine T
    - Larger interatomic distances between fold and rest of structure
    - Outer part of Patterson map

MR still "workhorse" to generate starting models for structural solutions in MX

Like Julius Caesar, however, the Patterson map is used to divide and conquer. The interatomic distances within the fold itself will be found in the Patterson maps within a radius approximately equal to the largest linear dimension of that fold, independent of where the fold sits within the unit cell. Hence, the first procedure is to determine the angular orientation R, which involves only three degrees of freedom. In our example above, this would require 0.1 seconds times 200 to the 2, which is about half an hour computer processing time. With this achieved, the outer region of the Patterson maps can be compared to determine the fold's position within the whole unit cell. Molecular replacement has been the workhorse for the starting models in most structural solutions in macromolecular crystallography. Only in de novo cases, have the earlier described approaches such as SAD been needed to be brought to bear.

Notes

Summary

6m 48s

In the next video...

In the next video of this section, we will discuss Alphafold 2, a machine learning algorithm that since its advent in late 2020, has revolutionised macromolecular structure determination.

Notes

Summary

7m 58s