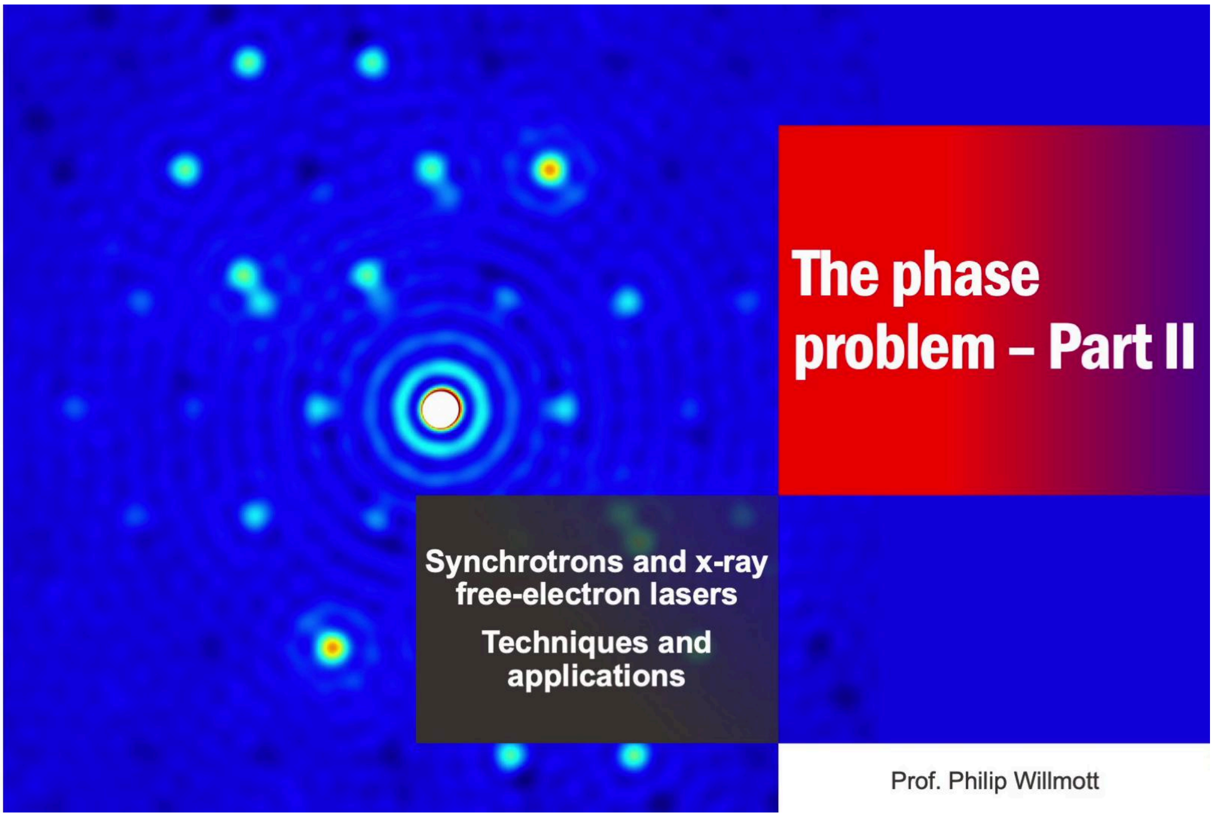
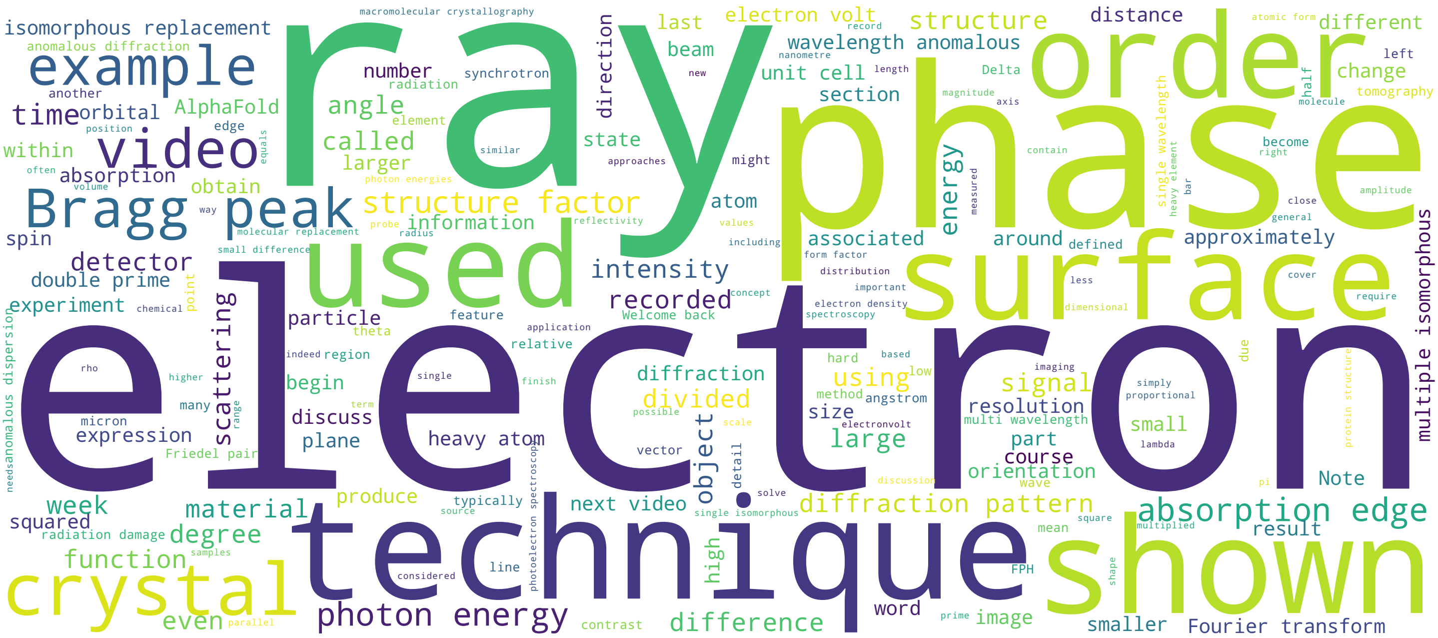
EPFL

École polytechnique fédérale de Lausanne





Search MOOC

Video

EPFL

Contents and objectives of this video

- Solving the phase problem for large molecules
- Different approaches
  - Multiple isomorphous replacement
  - Multiwavelength anomalous dispersion
  - Single-wavelength anomalous diffraction

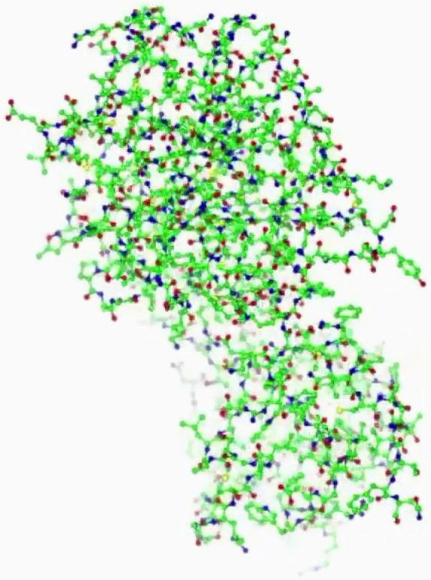
Welcome back. In this video, we discuss how one has solved in the past the phase problem for molecules that are so large that the approaches of direct and iterative methods presented thus far fail due to the complexity of the atomic basis. I will briefly cover the general approaches of the techniques multiple isomorphous replacement, multi-wavelength anomalous dispersion and single-wavelength anomalous diffraction, in this, their chronological order of development. In the next video, we look at the workhorse of phasing in macromolecular crystallography molecular replacement and the Patterson map on which this approach is based.

Notes

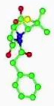
Summary

0m 05s

The problem



6LU7: COVID-19 main protease
2500 atoms (non-H), 34.5 kDa



3KP2: Benzylpenicillin
($C_{16}H_{18}N_2O_4S$)
23 atoms (non-H), 334 Da

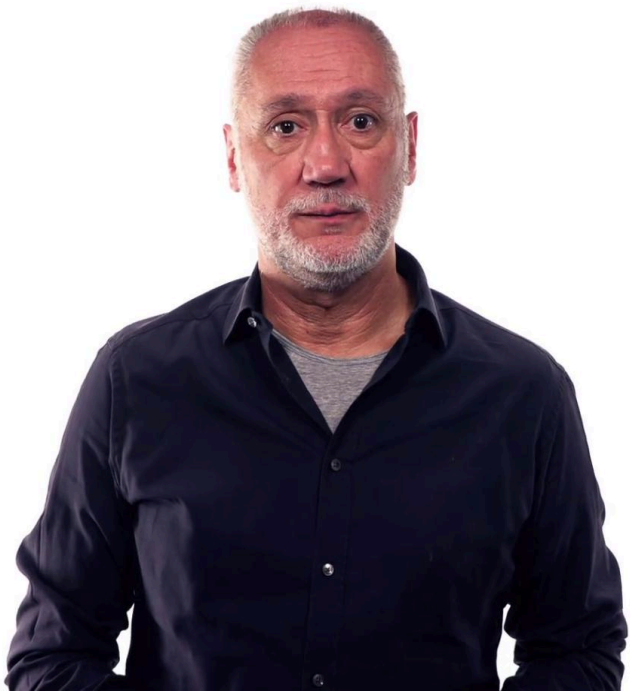
We can immediately see the problem here, even for the relatively modest protein structure of the COVID-19 main protease, which contains some 2,500 nonhydrogen atoms and has a molecular mass of approximately 35 kilodaltons, compared to benzylpenicillin, which is approximately 100 times smaller. Any attempt to use direct methods is bound to fail, as the typical atomic separations are so much smaller than the unit cell size. At least using even the most powerful modern computers, though who knows how this might develop in the future, iterative techniques also seem doomed to failure.

Notes

Summary

0m 52s

Macromolecular crystallography approaches



- Change something; observe differences; obtain structural information
- Multiple/single isomorphous replacement (MIR/SIR)
  - Change atomic/molecular component(s)
- Multiwavelength anomalous dispersion (MAD)
  - Change photon energy around an absorption edge
- Single-wavelength anomalous diffraction (SAD)
  - Observe differences in Friedel pairs
- Molecular replacement (MR)
  - Insert known molecular “chunk” as part of total structure
- AlphaFold2

In macromolecular phasing, the general approach has been, since the days of Perutz and Kendrew, that of changing something that does not significantly have an impact on the native structure, observe the differences that this change induces in the diffraction pattern and through this, obtain the necessary extra information to solve the phase problem. In the case of multiple and single isomorphous replacement, some part of the original molecular structure is substituted with another atomic or molecular chunk, normally chosen to be or contain a heavy element, which will significantly change the structure factors without distorting the original or native physical structure. This is the meaning of isomorphous; it means "same shape". A comparison between the native and isomorphously replaced structures provides the necessary information to solve the phase problem. Single isomorphous replacement is, as the name implies, a technique where only one modified structure is used in addition to the native structure. In multiple isomorphous replacement, two or more modified structures are used in order to overcome ambiguities associated with only using one substitute structure.

Notes

Summary

1m 40s

Macromolecular crystallography approaches

- Change something; observe differences; obtain structural information
- Multiple/single isomorphous replacement (MIR/SIR)
  - Change atomic/molecular component(s)
- Multiwavelength anomalous dispersion (MAD)
  - Change photon energy around an absorption edge
- Single-wavelength anomalous diffraction (SAD)
  - Observe differences in Friedel pairs
- Molecular replacement (MR)
  - Insert known molecular “chunk” as part of total structure
- AlphaFold2

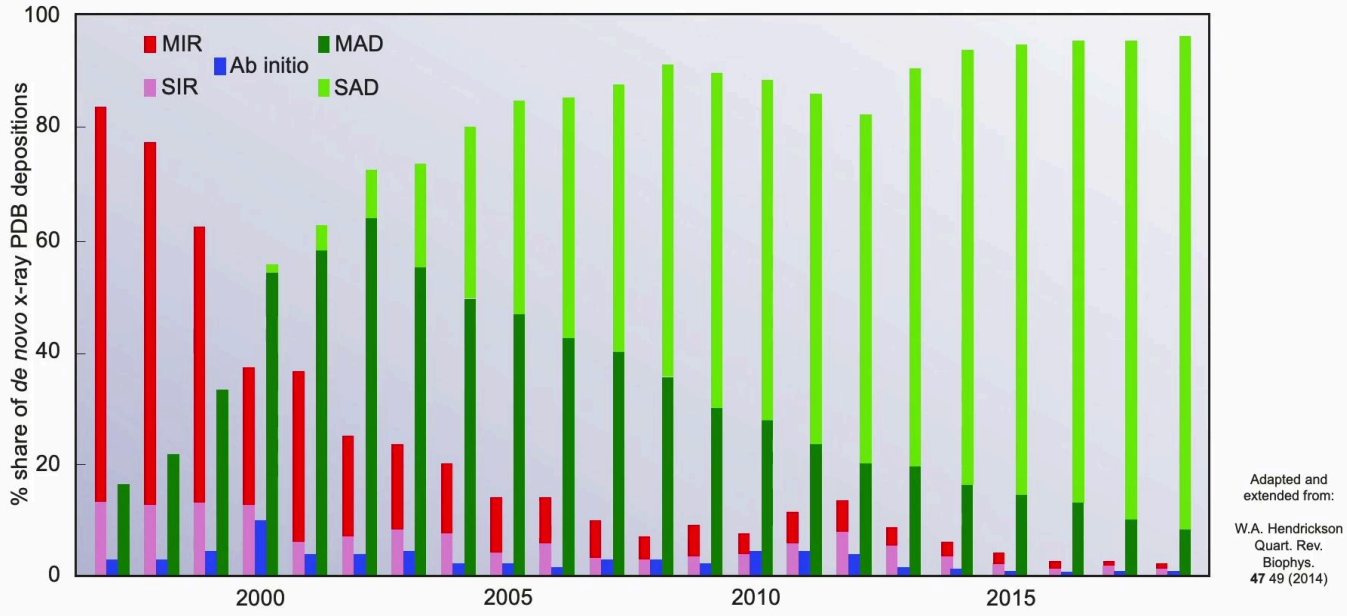
In multi-wavelength anomalous dispersion, the absorption edge of a heavy element substitution is more easily accessible to hard x-rays than the most common atomic components of biomolecules, namely hydrogen, carbon, oxygen and nitrogen. This is exploited in the differences in intensity of the Friedel pairs depending on whether the photon energy lies below, at, or above the absorption edge of the heavy element. In single-wavelength anomalous diffraction, only one photon energy is used somewhere above the absorption edge, normally. In the case of native SAD, the small difference in Friedal pair intensities is measured for elements such as phosphor or, more and most commonly, sulphur, which has its K-edge at approximately 2,500 electron volts. We will discuss these three approaches now. The method of molecular replacement, and most recently, artificial intelligence used in AlphaFold 2 are reserved to the last two videos of this section.

Notes

Summary

3m 12s

Trends in MX phasing techniques



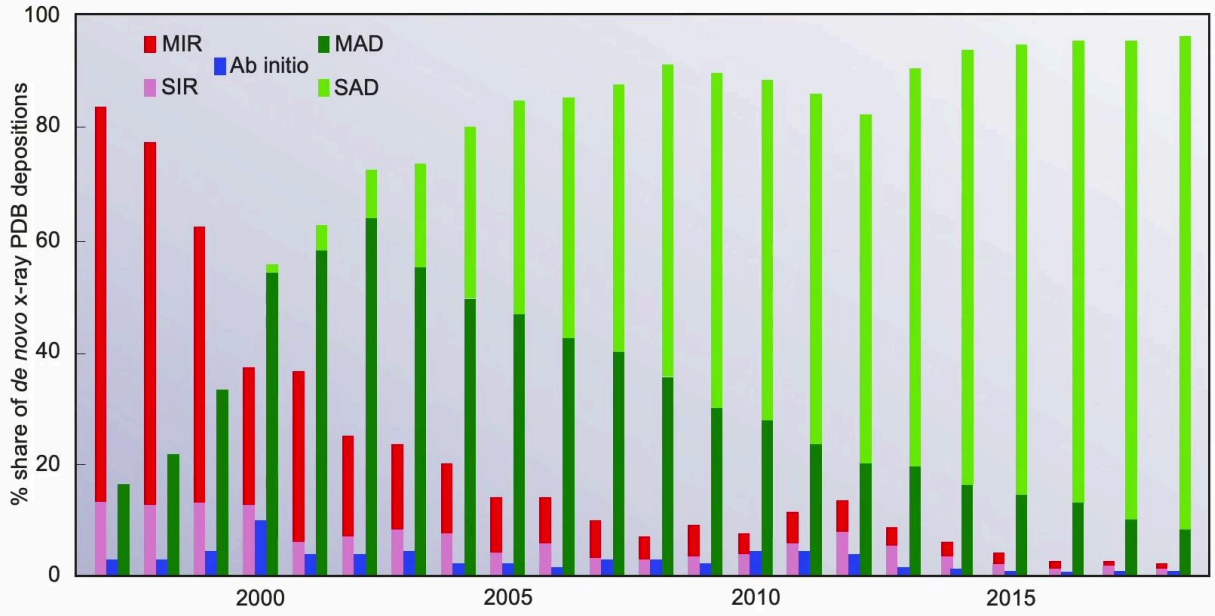
Before we continue, it's instructive to see the historical trends in phasing techniques in macromolecular crystallography. Just before the turn of the century, single and particularly multiple isomorphous replacement dominated phasing methods with small amounts of AB initio studies contributing too. None of these require tuning the photon energy to exploit anomalous effects. Now, once third-generation synchrotron facilities became established and mature, multi-wavelength anomalous dispersion took off, as in principle, one can record all three data sets using the same sample, although often radiation damage issues required the use of two or more crystals. Accompanying the developments in storage-ring performance were the similar drives in improvements in x-ray detector technology. Notably, after approximately 2005, with the introduction of hybrid photon counting area detectors such as the disruptive PILATUS detector technology developed at the Paul Scherrer Institute in Switzerland. The improved accuracy of these detectors and the capability of accurately determining the intensities of even the weak Bragg peaks meant that the small differences normally associated with single-wavelength anomalous diffraction could be accurately extracted.

Notes

Summary

4m 25s

Trends in MX phasing techniques



Adapted and extended from:

W.A. Hendrickson Quart. Rev. Biophys. **47** 49 (2014)

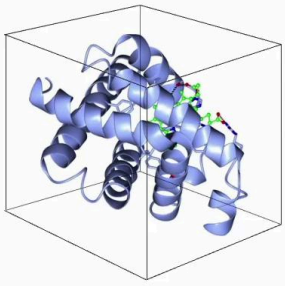
This was considered at the time, the nirvana of MX phasing, requiring, as it does, only one dataset. This method dominated phasing approaches until late 2020, when AlphaFold 2 was announced and shown to have, in a large fraction of cases anyway, to have solved the phase problem without the need for diffraction data. That said, recording diffraction data remains indispensable as a means of verification, which is absolutely essential in the world of pharmaceutical industries. But AlphaFold is reserved for the last video of this section, so let's not get ahead of ourselves.

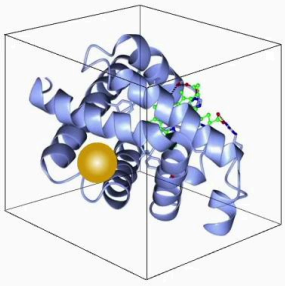
Notes

Summary

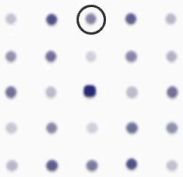
6m 00s

Multiple/single isomorphous replacement

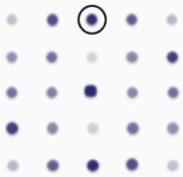


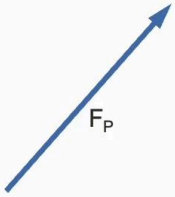


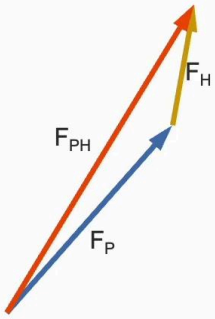
$|F_P|^2$



$|F_{PH}|^2$







- Requires macromolecule to accept heavy atom and not change its shape ("isomorphous")
  - Substitution e.g. Se for S
  - Addition e.g. Hg^{2+} added
- Unambiguous information only if two different "heavy-atom substitute" crystals can be created (MIR)
- Use "Harker diagrams"
  - See description in "Introduction to synchrotron radiation", Section 6.11.4

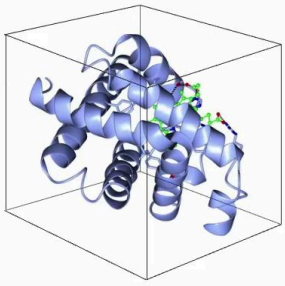
Because the phasing techniques prior to the advent of AlphaFold 2 are now largely redundant, I will only cover them here very superficially. For a more in-depth discussion, I refer you to my textbook, Section 6.11.4. We begin here with multiple and single isomorphous replacement. A crystal composed of native proteins, that is, proteins without any isomorphous additions or substitutions produces a certain set of structure-factor intensities FP squared, in other words, a diffraction pattern. A second crystal is synthesised with the same shape, but with part of it substituted or added containing a heavy atom. This produces its own diffraction pattern composed of structure factors FPH, where the H donates the heavy component. Now let's focus on one Bragg peak, which arises due to the vector addition of all the atomic form factors of the constituent atoms to produce FP. Now, the same Bragg peak in the modified protein structure has the same structure factor as the native version, plus an extra component FH, to produce FPH. One of the most common substitutions is that of replacing sulphur with its atomic number 16 with the chemically very similar selenium.

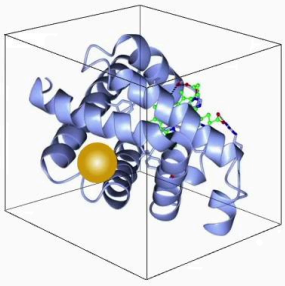
Notes

Summary

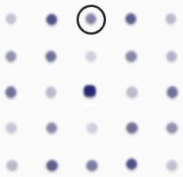
6m 40s

Multiple/single isomorphous replacement

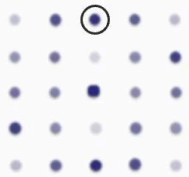


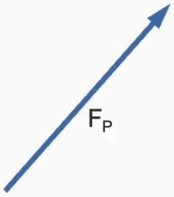


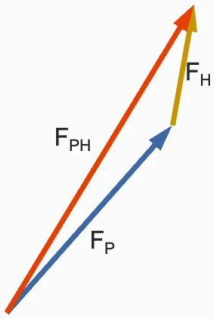
$|F_P|^2$



$|F_{PH}|^2$







- Requires macromolecule to accept heavy atom and not change its shape ("isomorphous")
  - Substitution e.g. Se for S
  - Addition e.g. Hg^{2+} added
- Unambiguous information only if two different "heavy-atom substitute" crystals can be created (MIR)
- Use "Harker diagrams"
  - See description in "Introduction to synchrotron radiation", Section 6.11.4

Max Perutz, haemoglobin, 1959
Chemistry Nobel Prize 1962

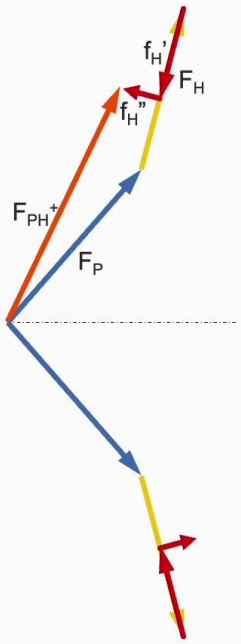
Selenium lies directly below sulphur in the periodic table, which has an atomic number of 34. Other additive species include, for example, mercury ions. Note that without some other trick, unambiguous information can only be extracted if there are two or more heavy-atom-derivative crystals. One of the major stumbling blocks was to synthesise crystals that would adopt these substitutes at all, never mind in an isomorphous fashion. The unambiguous phases are extracted thanks to so-called Harker diagrams, also explained in more detail in my textbook. It was Max Perutz who pioneered this technique in his gargantuan effort to solve the structure of haemoglobin. For this labour of love, he would receive the Nobel Prize in Chemistry along with John Kendrew for his studies on myoglobin in 1962.

Notes

Summary

8m 16s

Multiwavelength anomalous dispersion



- Single molecular structure
  - Includes heavy atom, e.g. Se
- Record three or more diffraction sets at energies around a strong absorption edge of a heavy element H present in macromolecule
  - Only possible at synchrotrons
- $F_{PH} = F_{PH}(h\nu, \text{Friedel pair})$
  - $|F_{PH}(h\nu_1)| \neq |F_{PH}(h\nu_2)|$
  - $|F_{PH}^+| \neq |F_{PH}^-|$
- MAD = “perfect MIR experiment”: isomorphism guaranteed!
- Reliability of analysis very sensitive to radiation damage

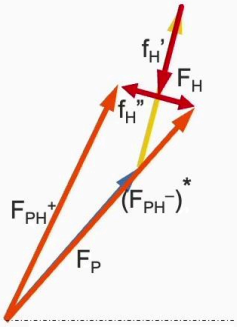
Multi-wavelength anomalous dispersion has much in common with multiple isomorphous replacement, insofar that three or more diffraction sets are recorded with differences in their intensities enabling the phases to be extracted. The difference here is that this is done with a single molecular structure, one that, in most cases, already has a heavy atom inserted into it. The changes are induced not by changing the heavy atom, as is the case in multiple isomorphous replacement, but instead by changing the photon energy around an absorption edge of the heavy-atom substitute. As such, one can think that MAD is a perfect MIR experiment, with isomorphism guaranteed, though one needs to be aware of artefacts arising due to radiation damage. In detail, MAD works as follows: the structure factor of the native protein FP is vectorially added to the atomic form factor of the heavy atom FH. But because we are near the absorption edge of the heavy atom, we need to consider f prime and f double prime in order to obtain our structure factor, which we label FPH plus. We now consider the other structure factor of the Friedel pair, which mirrors that of FPH plus, except that f double prime points in the same sense relative to f prime.

Notes

Summary

9m 21s

Multiwavelength anomalous dispersion



- Single molecular structure
  - Includes heavy atom, e.g. Se
- Record three or more diffraction sets at energies around a strong absorption edge of a heavy element H present in macromolecule
  - Only possible at synchrotrons
- $F_{PH} = F_{PH}(h\nu, \text{Friedel pair})$
  - $|F_{PH}(h\nu_1)| \neq |F_{PH}(h\nu_2)|$
  - $|F_{PH}^+| \neq |F_{PH}^-|$
- MAD = “perfect MIR experiment”: isomorphism guaranteed!
- Reliability of analysis very sensitive to radiation damage

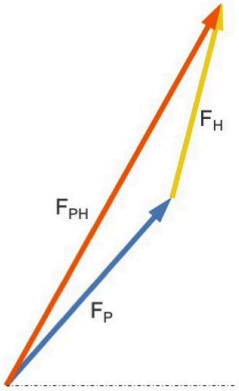
To make our task of seeing the differences between FPH plus and FPH minus, we flip the latter across the real axis so we can see the difference between them more clearly. The asterisk is shown to indicate that as a complex number in the Argand diagram, the flipped set of vectors is the complex conjugate of FPH minus.

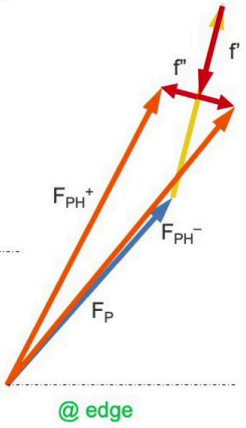
Notes

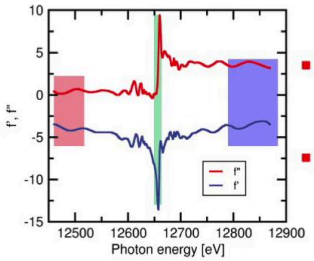
Summary

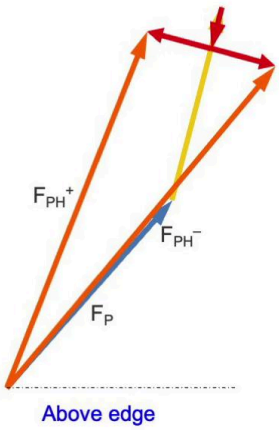
11m 01s

Multiwavelength anomalous dispersion









- Single molecular structure
  - Includes heavy atom, e.g. Se
- Record three or more diffraction sets at energies around a strong absorption edge of a heavy element H present in macromolecule
  - Only possible at synchrotrons
- $F_{PH} = F_{PH}(h\nu, \text{Friedel pair})$
  - $|F_{PH}(h\nu_1)| \neq |F_{PH}(h\nu_2)|$
  - $|F_{PH}^+| \neq |F_{PH}^-|$
- MAD = “perfect MIR experiment”: isomorphism guaranteed!
- Reliability of analysis very sensitive to radiation damage

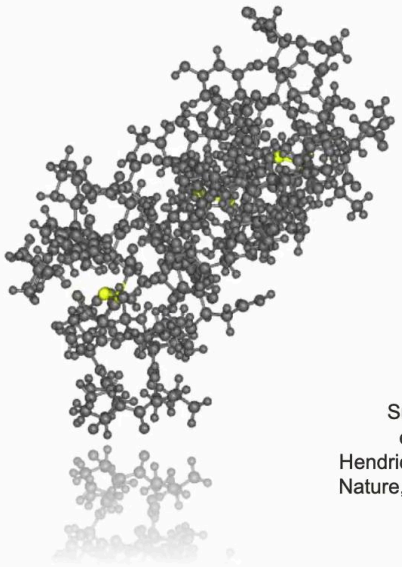
In summary, well below an absorption edge, a standard diffraction pattern will be recorded in which there are only extremely small differences in the intensities of Friedel pairs, if any at all. At the absorption edge, f prime can be large, as can f double prime, depending exactly where the photon energy lies on the absorption curve. At still higher photon energies, f prime will shrink again substantially, though f double prime can persist for a few hundred electron volts. Exactly how f prime and f double prime impact the intensity of the Bragg peak depends on the orientation, in other words, the phase of FH relative to FP.

Notes

Summary

11m 30s

Single-wavelength anomalous diffraction



Sulphur-SAD
of crambin
Hendrickson and Teeter
Nature, **290**, 107 (1981)

- One data set only
  - Method with most rapid data acquisition
  - Minimum radiation damage
- On one crystal type
  - Heavy-atom derivative
  - Naturally occurring element (native SAD)
- Exploits difference in Bijvoet pairs F^+ and F^-
- Most demanding technique regarding accuracy of recording structure factors

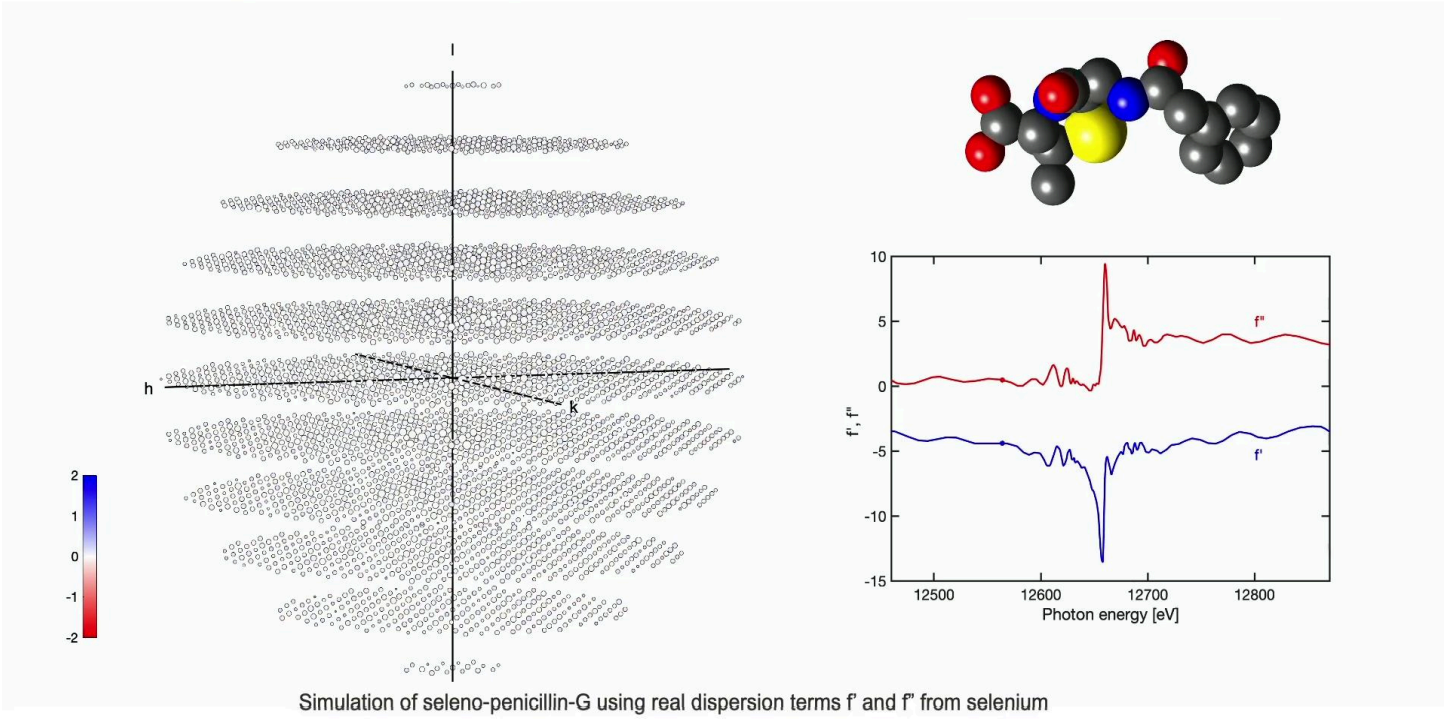
The last technique we discuss in this video is single-wavelength anomalous diffraction. In this technique, only one dataset is recorded and is thus the fastest regarding data acquisition and with the least associated radiation damage. The crystal may contain a heavy atom such as selenium, but more recently native SAD has become the first technique of choice, where the small differences in structure factors in the Friedel or Bijvoet pairs due to anomalous effects in naturally occurring elements such as sulphur are recorded. The differences are larger, the closer one is to the absorbing atom, but has associated disadvantages too. If one goes too low in the photon energy, radiation damage increases, absorption by the sample environment, such as air increases, and the accessible volume of reciprocal space, and consequently the number of structure factors also decreases. Thus, there is normally a compromise between getting a large anomalous difference signal and the ease of experimentation and structural resolution.

Notes

Summary

12m 16s

Single-wavelength anomalous diffraction



Simulation of seleno-penicillin-G using real dispersion terms f' and f" from selenium

Shown here, is a simulation of the relative differences in Friedel pair intensities as a function of photon energy for the derivative structure of seleno-benzylpenicillin. We can expect the largest deviations for the weakest diffraction maxima, as it is for these that the vector addition of f prime and f double prime of selenium will make the most relative difference. The size of the circles representing each Bragg peak indicates the peak's intensities. Note also that well below the absorption edge at 12,664 electron volts, the SAD signal is extremely weak, but then becomes strongest right around the absorption edge and persisting all the way up to and beyond the highest shown photon energy of 12,860 electron volts, thanks in large part to f double prime remaining significantly non-zero.

Notes

Summary

13m 28s

In the next video, we will look at molecular replacement, a technique that has profited and improved as a database for protein structures has grown over the last three decades.

Notes

Summary

14m 29s