5-2003

# Protein Crystallography Retinol Dehydratase Y135F Mutant Project Expression, Purification, and Crystallization

Samuel Costa Araujo

# Protein Crystallography
# Retinol Dehydratase Y135F Mutant Project
# Expression, Purification, and Crystallization

by
Samuel Araujo

Undergraduate honors thesis under the direction of
Dr. Marcia Newcomer
Department of Biological Sciences

Submitted to the LSU Honors College in partial fulfillment of
the Upper Division Honors Program

May 2003

Louisiana State University

**Introduction.**

The discovery of the diffraction of X-rays by crystals is owed to Max von Laue in 1912 (1). In spite of the fact that Wilhelm Roentgen discovered X-rays in 1885 (almost 30 years prior to Laue's discovery), their nature was then still unknown (1). Laue's discovery was paramount in proving the wave character of X-rays (1). Moreover, he was awarded a Nobel Prize in Physics in 1914 "for his discovery of the diffraction of X-rays by crystals." In the same year that Laue diffracted X-rays by crystals, the Bragg brothers derived Bragg's Law of Diffraction, which is a relationship between diffraction patterns and the relative position of the diffracting points.

X-ray crystallography has revolutionized the understanding of protein structure and their function by allowing us to look into the three dimensional location of nearly each atom in a protein (2). This structure-defining technique unveiled the structure of the first globular protein ever to have its three-dimensional configuration revealed. This protein was myoglobin (3).

Through X-ray crystallography we are able to learn how certain proteins fold. By understanding how the secondary structures interact in order to shape the tertiary and quaternary structures, and ultimately the entire protein, it is possible to theorize on the functions of a protein, as well as how it interacts with possible substrates and cofactors.

R etroretinoids are vitamin A derivatives that have a shift in their conjugated double bonds in such way that the bond that joins the isoprene tail to the ionone ring is a double bond (4). These derivatives of dietary retinol are an extremely important requirement in a variety of biological processes such as spermatogenesis and the maintenance of epithelial tissue and vision (5).

Retinol dehydratase (Fig. 1) from *Spodoptera frugiperda* is a part of the superfamily of enzymes called sulfotransferases (6). It is allegedly involved with the dehydration process of retinol via an unstable intermediate of retinyl sulfate (7).

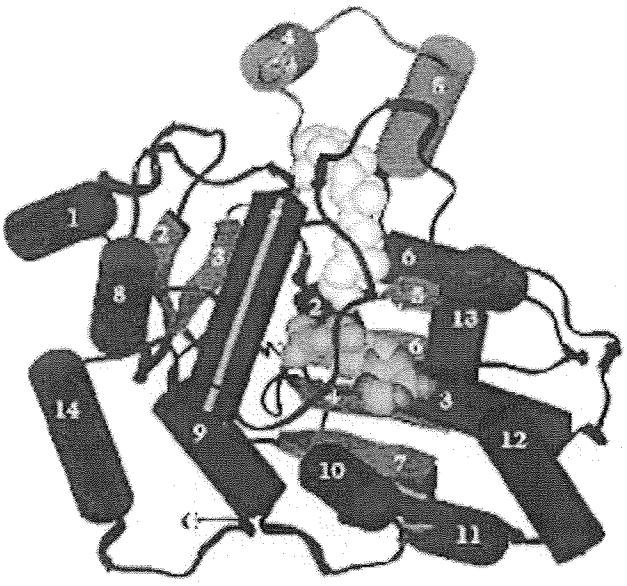


Fig. 1. Retinol dehydratase with retinol in its active site
From Pakhomova *et al.* (2001), p. 448.

I worked with a mutant of retinol dehydratase. This mutant has one of its Tyrosines replaced by a Phenylalanine in the 135 position (Fig. 2). This particular mutation is interesting because, although it represents a small change in the protein's molecular weight and overall composition, it completely inactivates it. The Y135F retinol dehydratase mutant was reported to have a molecular weight of 41.4865 kDa by the ProtParam online tool.

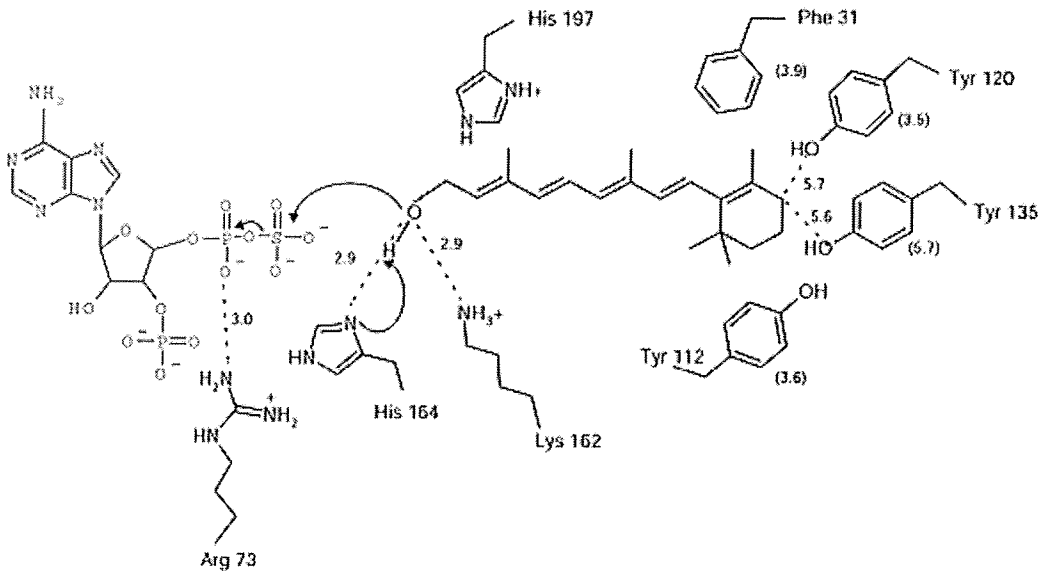


Fig. 2. Retinol dehydratase's active site showing the amino acid that will be substituted (Tyr 135) as well as its interaction with retinol. From Pakhomova *et al.* (2001), p. 449.

The process of finding a protein's three-dimensional structure encompasses many steps. In this paper I will deliver an overview of what has been accomplished in this project, and report my basic understanding of the science behind the future of this project.

**A. Transformation.**

*E. coli* bacteria are transformed using plasmids provided by a collaborator. The many features of these plasmids included an ampicillin-resistance gene, the DHR Y135F mutant gene, and the T7 promoter adjacent to it. The strain of *E. coli* bacteria used was BL21(DE3), a chemically competent strain, which means that it had its membranes chemically treated so that pores could be formed on it, thus allowing these cells to take up DNA. This strain also had the DE3 region of its genomic DNA engineered so that it contained the gene coding for the T7 RNA polymerase, which is

under the control of a chemical called IPTG (iso-propyl-thio-β-D-galactopyranside). The heat shock method was used in order to drive the plasmids into the bacteria.

**B. Expression/Harvesting.**

The bacteria were first cultured in liquid media, so that they would have time to develop ampicillin resistance before being streaked onto the ampicillin-LB plates. After considerable growth was observed, the bacteria were streaked onto ampicillin-LB plates, which selected for our ampicillin-resistant bacteria. Once colonies were spotted, a single flask containing liquid broth was inoculated with the bacteria and allowed to grow overnight.

The next day, several flasks were inoculated with our bacteria. We wanted them to be in the log phase of their life cycle in order to add IPTG to the broth. The way this was monitored was by checking the optical density at 600nm of the broth over time, using the spectrophotometer. Once we obtained an O.D. 600 value between 0.6 and 0.8, we knew they had reached log phase, at which point IPTG was added.

In general, IPTG (iso-propyl-thio-β-D-galactopyranside) is used to induce the bacteria to produce the protein of interest in a larger quantity than it normally would. As a matter of fact, this target protein is produced almost exclusively. The way this mechanism works is related to the fact that this particular strain of *E. coli* contains the T7 RNA polymerase gene that is dependant on the presence of IPTG to be turned on. The reason for this is that IPTG binds the *lac* repressor and remove it from its repressing position in the genomic DNA. Once this gene was turned on, the bacteria produced T7 RNA polymerase, which, in turn, binded to the T7 promoter in our DHR Y135F plasmid, thus producing our protein. The bacteria were incubated for about three hours after the

IPTG induction. The broth was then centrifuged at 5,000 g, and the pellets that formed were frozen.

**C. Purification**

As we know, purity is the key to the formation of viable crystals, which will have its proteins as packed as possible, thus allowing for less room between molecules where water can lodge itself. Consequently, the purer the protein is, the better the diffraction patterns obtained. The purification process was based on that of a previously published work by Vakiani *et al.* (7).

The frozen cell pellets from the harvesting step were thawed, which begins the rupturing of the cell membranes. The cells were then resuspended in DHR lysis buffer. This lysis buffer had a pH of 7.4 (which is the physiological pH), NaCl (for ion strength, as some proteins like to be in the presence of salts), DTT (reducing agent that makes sure that all disulfide bonds are kept intact: none are broken, none are formed), and EDTA (a metal chelator to make sure that all metals present are isolated, thus impairing all metal dependent proteases).

The resuspended cells are then sonicated, which breaks the already ruptured membranes fully apart, as well as breaks the DNA. An evidence of DNA breakage is the loss of viscosity of the solution as it is sonicated. After the process of sonication is complete, we have what is called cell lysate.

The cell lysate contains insoluble matter (cell debris), soluble proteins, soluble molecules from *E. coli*, and broken DNA. The lysate was spun down at 40,000 g in order to separate the insoluble matter from the soluble phase. After this process, the supernatant was collected and the pellets discarded.

As you may recall, our protein was last resuspended in a buffer containing NaCl. After the cell lysate was centrifuged so that the cell debris could be separated from the proteins in solution, it was dialyzed so that we could get rid of the NaCl in the lysis buffer. The cell lysate was placed in a dialysis bag, which contains pores large enough to let the sodium and chloride ions diffuse out along with other small impurities. The dialysis bag was placed in low salt buffer overnight.

Following dialysis, the protein was eluted through a DE52 column. This column contains a positively charged resin, which attracts my negatively charged protein. The gradual increase of salinity in the buffer running through this column will make different proteins come off at different times. Since fractions were collected and the molecular weight of my protein was known, we were able to run SDS-page gels after this procedure in order to identify the fractions in which my protein was (Fig. 3). The gel shown in figure 3 represents every fourth fraction collected from the DE52 column. More inclusive gels (every other fraction) were run later in order to better identify the fractions that contained my protein.

The gel shows the fractions containing the highest concentration of my protein after it ran through this column. The fractions containing the highest concentration of my protein with the least amount of impurities were pooled, and all others were discarded. Although this picture lacked quality and was a preliminary one, it showed the different levels of impurities in each lane.

A B

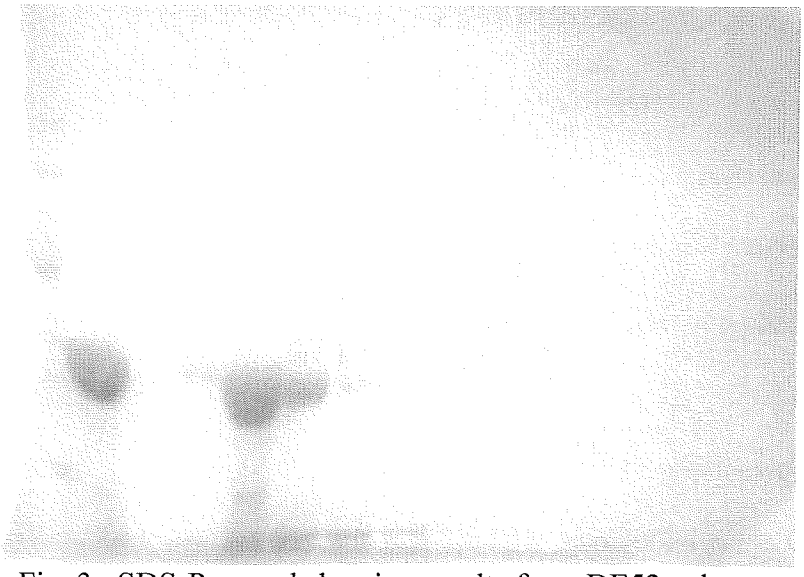

Fig. 3. SDS-Page gel showing results from DE52 column. Lane "A" shows the molecular marker, lane "B" shows a completely purified sample of retinol dehydratase. All other lanes show fractions collected from column.

Following the previous step, the fractions that were pooled were concentrated and then eluted through a size-exclusion column. This type of column is made of porous, microscopic beads. The sizes of these pores determine when proteins will come off the column. If a protein is small enough to fit in them, they will stay in the column longer than larger proteins that will not fit, and consequently go right through the column. This way we were able to separate my protein from all others based on size. Once again, fractions were collected, and then run through SDS-page gels in order to verify the purity of my protein at this point as well as which fractions contained it (Fig. 4). We were surprised to notice that not only was the protein rather pure for this step, but also that most fractions contained a high concentration of it. Considering that we were pretty sure of what the gel would look like, as far as finding our protein, no molecular

markers or purified proteins were ran in it. Given the purity at this point, all fractions were collected, pooled, and carried to the next step.

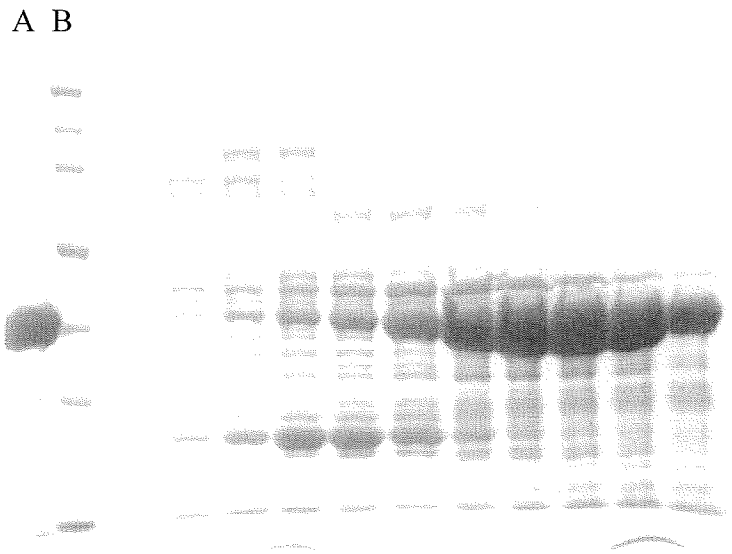


Fig. 4. SDS-Page gel showing results from size-exclusion column, every other fraction. Lane "B" shows the molecular marker, lane "A" shows a completely purified sample of retinol dehydratase. All other lanes show fractions collected from column.

Normally, the fractions collected from the size-exclusion column would be concentrated prior to the next step. However, due to the already high concentration of my protein in each individual fraction and the high number of fractions, we simply had to run the protein solution through the next column several times.

The next column in the process is called a mono-Q column, which is assembled in a Forced Pressure Liquid Chromatograph (FPLC). The mono-Q column basically works the same way as DE52 does (positively charged resin, negatively charged protein, change is salt concentration in buffer through column). However, the latter is cruder than the former, therefore it is used in the initial step of purification, at which point the protein solution is full of impurities. Through the mono-Q column, we ran low

salt buffer at first, just so impurities that do not bind to the column would flow right through. The gradual increase in salt concentration allowed different things to flow through at different times.

The fractions collected from the FPLC machine were believed to be rather pure fractions containing our protein. The reason for this assumption is that we were able to measure high absorbance peaks at 280 nm for these fractions at a point in the salt gradient that is the expected for this protein. SDS-page gels were run in order to confirm purity (Fig. 5).

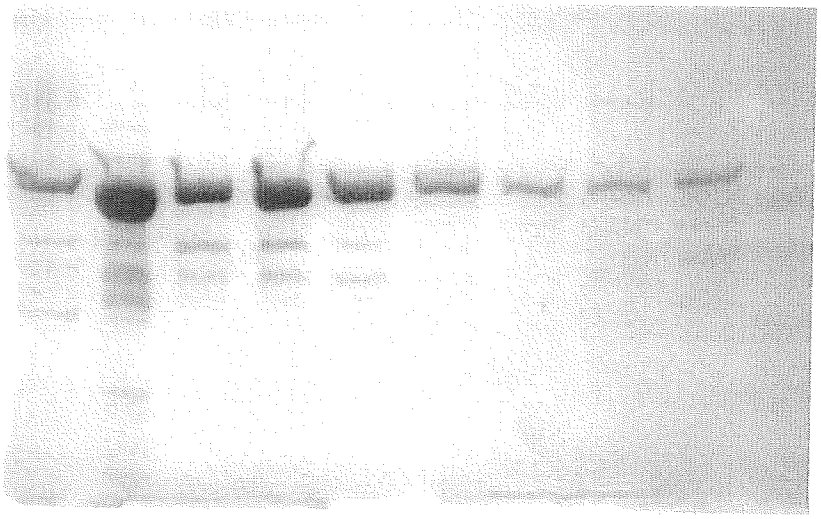

Fig. 5. SDS-Page gel showing results from Mono-Q column, every other fraction. No molecular weight marker was used, given that at this point we were certain we had our protein of interest.

**D. Crystallization**

Inorganic substances can easily yield crystals when in a hot, saturated solution of this substance and then slowly cooled down (8). This method can sometimes

be used to obtain crystals from polar organic substances, along with the controlled addition of organic solvents (8).

However, as common sense calls, these conditions are not adequate for the proper handling of proteins to be crystallized, considering that they easily denature when exposed to heat or organic solvents.

When a protein is more than 97% pure according to standard criteria of homogeneity (3) it can be crystallized by dissolving it in an aqueous buffer containing a precipitant, usually ammonium sulfate or PEG (polyethylene glycol), at a concentration that is barely below the threshold for precipitation (8). The water is then removed from this precipitation solution through a controlled process of evaporation (3), which produces precipitation conditions.

The hanging-drop method is one of the extensively used techniques for this process (8). The hanging-drop method consists of using a glass cover slip to cover a precipitant-containing well, and vacuum grease is used to seal the assembly (Fig. 6). Before covering the well with the glass cover slip, a drop of the well solution (precipitate) is placed on the glass cover slip surface that will face the inside of the assembly, followed by the addition of no more than 25 μL (5) of protein to this drop. The amount of protein added to the drop is generally the same volume as the drop, which gives a precipitant concentration approximately 50% of that needed for precipitation. Since the precipitant is the major solute in the closed assembly, transfer of water from the protein-containing drop to the well solution is the overall outcome of vapor diffusion (8). This process continues until equilibrium is reached. Usually, these wells are arranged in trays, which are called crystal trays (like, what a shocker!!!).

If all goes well, crystals should form (Fig. 7). One expects to get few crystals that are large enough for diffraction. However, it may happen that many crystals are formed, which will consume the entire protein in the drop before the crystals are large enough to be of any use.

In the event that small crystals of good quality are formed, they can be used as seeds to grow other crystals, which, in general, grow faster than the ones that grow straight from the drop-protein solution in the hanging-drop assembly (8). The arrangement of the hanging-drop assembly for seeding is the same as the one described above.

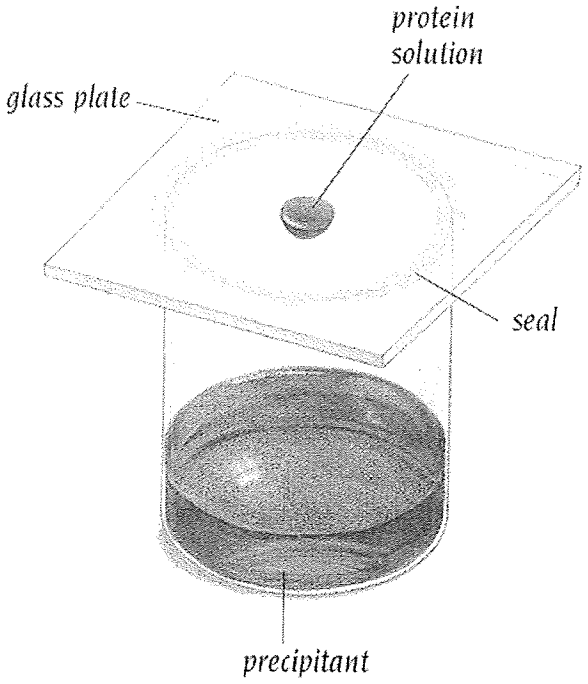


Fig. 6. The hanging-drop method.
From Branden and Tooze, p. 375.

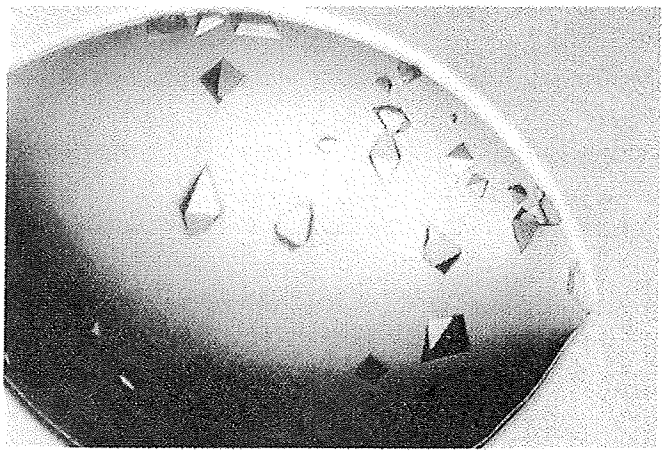

Fig. 7. Example of hanging drop with crystals.
From Branden and Tooze, p. 375.

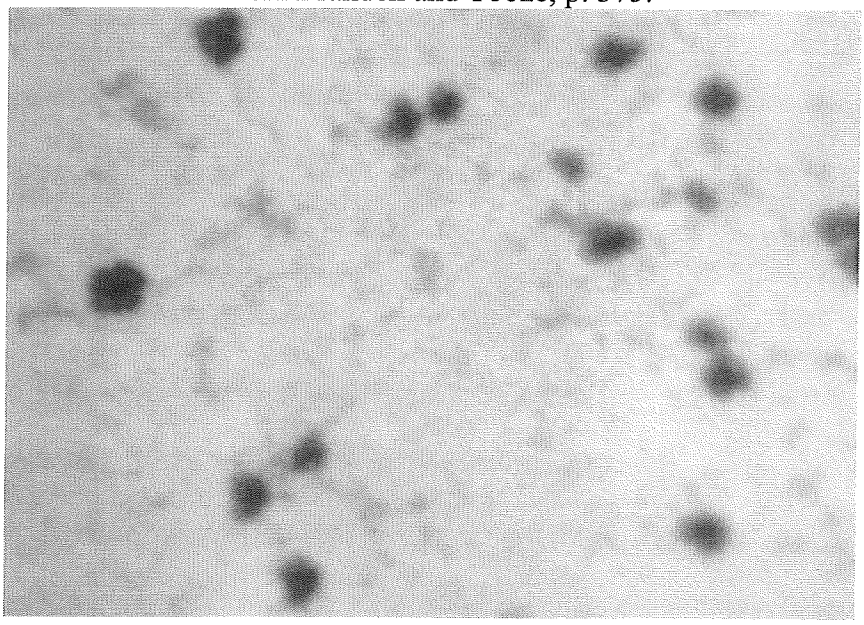

Fig. 8. Spherulites from well with Index Solution 85.

**E. Diffraction**

Viable crystals are necessary for high-quality diffraction. Large, symmetrical crystals do not necessarily equate viability. As a matter of fact, better diffraction can, sometimes, be obtained from cracked, asymmetric looking crystals than from neatly symmetric ones. In general, the closer the protein molecules are packed, and

consequently the less water they trap among themselves, the better is the diffraction pattern generated by them (3). The reason for this is that the closer they are packed in a crystal, the better ordered they are.

The wavelength of X-rays is precise enough to detect atoms separated by the distance of a covalent bond (6). In addition, the energy of a quantum of the X-ray radiation is roughly 8000 eV, which coincides with the approximate energy of electrons in their orbitals (9). These properties make X-rays more than well suited for the study of protein structure through the use of diffraction.

Diffraction is determined by how much interference the X-rays are exposed to. The electron density around an atom, which is the number of electrons in a determined space, is a direct determinant of how strongly X-rays are scattered by this particular atom (8). Thus, the diffraction patterns generated by crystals are a direct clue to the position of each atom in the target protein.

An apparatus called diffractometer is used in the collection of diffraction data from protein crystals. First, the crystals are harvested from the crystal trays using a small (approximately 0.5 mm in diameter) loop, which is then attached to a pivoting axis on the diffractometer. This pivoting axis allows the user to calibrate its rotation so that the X-ray beam will hit the crystal at all times during the crystal's rotation. The X-ray source is placed at a 90° angle to the axis of rotation of the crystal. Directly on the other side of the crystal, and perpendicular to the place of the X-ray beam, is the detector, which can be placed at different distances from the crystal. The detection surface of the detector is placed perpendicular to the X-ray beam, which places it in a separate plane from the other parts of the diffractometer.

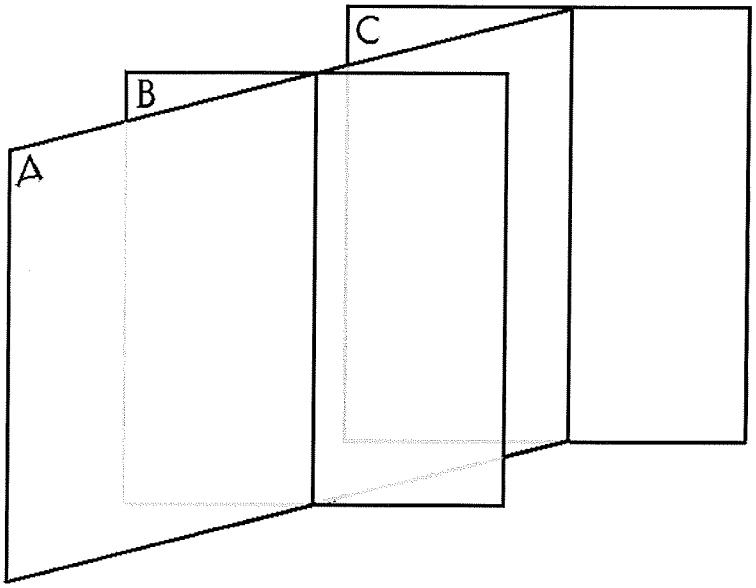


Fig. 9. Diffractometer setup.

In short, most diffractometer apparatuses are placed in three different planes (Fig. 9), two being parallel: C (the one containing the detector) and B (the one which contains the crystal pivoting axis and the liquid nitrogen stream), and a third one which is perpendicular to the previous two: A (which contains the X-ray beam).

One of the problems encountered by crystallographers is that crystals, like the proteins that make them, are rather unstable when exposed to heat and other extreme conditions. Moreover, crystals tend to suffer deterioration when hit by the X-ray beam, which is directly related to free radicals that are generated by the X-rays (8). To minimize this adverse effect, the protein crystals are kept frozen with liquid nitrogen. Directly pointed to the crystal, and in the same plane, is the liquid nitrogen stream that flows from the cryo-cooling apparatus of the diffractometer.

The data generated and collected by the diffractometer give us ground upon which we can analyze the way the protein crystal being studied can change the direction of the X-ray radiation waves. In order to better understand the scattering of X-

rays and how it helps us discover the three-dimensional structure of a protein, let us consider the example of a mirror. A mirror simply changes the direction of light waves by reflecting them. By the same token, nonreflecting objects can also change the direction of light waves by scattering it. This explains why objects placed in the path of a light beam can never cast a perfectly sharp shadow, for the edges of this object are scattering the light waves. This phenomenon can be explained by Huygen's Principle, which says that every point in a wavefront can be considered a new origin of wavefronts (9) (Fig. 10.). These new wavefronts are called scattered waves.

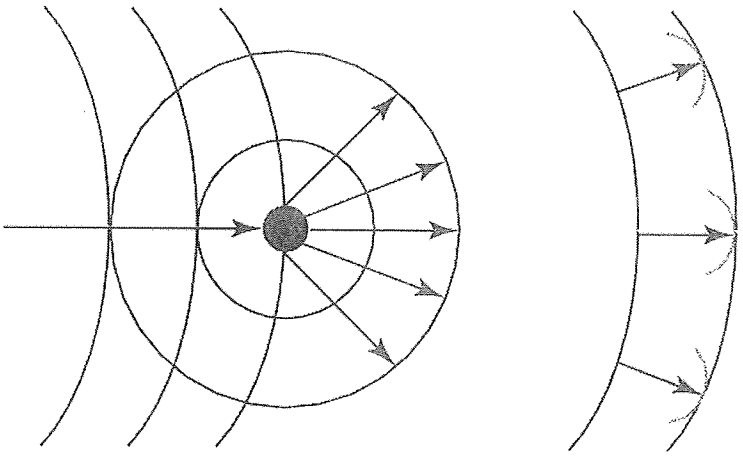

Fig. 10. Huygen's Principle.
From van Holde *et al.*, p. 259.

Now let us consider two point objects, A and B, placed in front of a wavefront (Fig. 11 (a)). These two point objects will produce scattered waves upon the passing of the original wavefront. The outcome of the interactions between these two new wavefronts will determine the position and consequently the amplitude of them. For instance, the amplitude will be doubled at points in which the scattered waves are perfectly in phase, and will equal zero when the scattered waves are perfectly out of

phase (Fig. 11 (b) and (c)). The information collected from the points at which the scattered waves interact (such as location, amplitude, and distance from each other) is directly related to the distance between the point objects A and B as well as to their distance from the source of the original front.

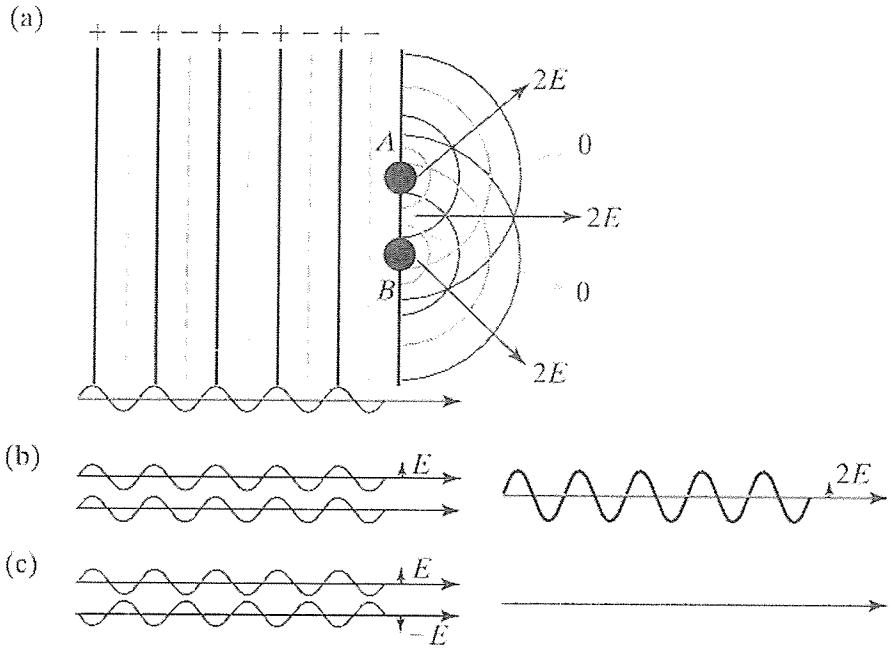


Fig. 11. Constructive vs. destructive interference.
From van Holde *et al.*, p. 260.

Considering that X-rays behave the same way the waves described above do, their diffraction will generate scattered waves that will, in turn, interact with one another. The data generated by these interactions is recorded by the diffractometer's detector and ultimately used to determine the position of the "diffracting objects" (atoms in a protein).

In the same year that Laue discovered that crystals diffracted X-rays (1912), William L. Bragg along with his brother William H. Bragg derived a simple relationship to understand how diffraction is related to the relative position of point

objects in space (9). In order to derive Bragg's Law of Diffraction, we first have to think of the lattice points in the crystal as parallel planes (Fig. 12). Now if we pile these planes up with an equal distance $d$ between them, thus creating a one-dimensional crystal model, and then shine X-ray radiation of wavelength $\lambda$ at a $\theta$ angle of incidence, we can start to wonder which values of θ will result in constructive or destructive interference. For this model, we assume that the distance $d$ is infinitely smaller if compared to the distance between the planes and our point of observation. Thus, the individual paths of the scattered X-ray beams are fundamentally parallel (9).

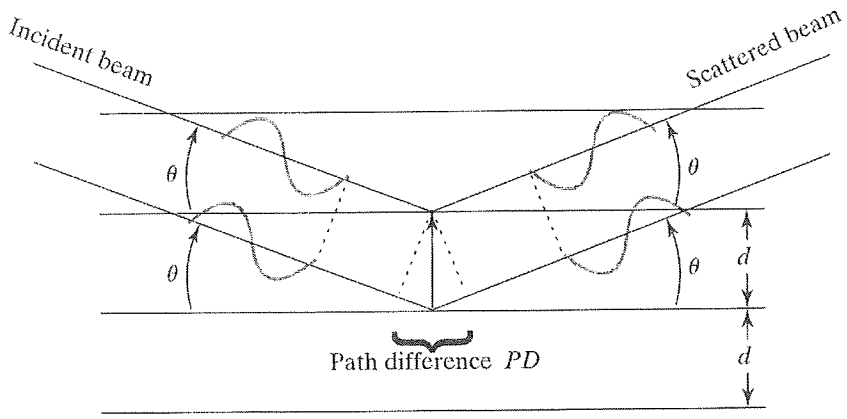


Fig. 12. Bragg's law model.
From van Holde *et al.*, p. 261.

Due to the large number of planes in this model, only when the scattered waves are perfectly in phase can constructive interference truly occur. For these waves to be in phase, the difference in the path length of the incident and the reflected waves of each plane, $PD$, is equal to an integer $n$ of the wavelength of the incident X-rays (9). This give us $PD = n\lambda$. A simple trigonometric relationship relates $PD$ to the distance $d$

between the planes in this model: $\frac{1}{2}PD = d \sin \theta$. If we combine these two equations we derive Bragg's Law of Diffraction: $2d \sin \theta = n\lambda$ or $2 \sin \theta = n\lambda/d$.

All this work earned the Bragg brothers a Nobel Prize in Physics in 1915 "for their services in the analysis of crystal structure by means of X-rays."

**F. Phase determination**

In light microscopy, lenses focus the light beams that are scattered by the observed object. This way, the image is directly formed from the scattered beams of light. X-rays, however, cannot be focused in the same manner.

In order to understand the patterns created by the scattered X-ray beams, and consequently "see" the structure of the proteins scattering the beams, we employ a mathematical relation called a Fourier transform (2).

Waves, such as X-ray radiation waves, can be expressed as periodic functions, and simple wave equations can be written in the form $f(x) = F\cos 2\pi(hx + \alpha)$ or $f(x) = F\sin 2\pi(hx + \alpha)$, where $f(x)$ is the vertical height of the wave at a given horizontal position (in wavelengths, where $x = 1$ represents a full cycle in the periodic function), $F$ is the amplitude of the wave, $h$ is its frequency, and $\alpha$ specifies its phase (position in relation to the origin) (8).

Any complex wave can be expressed as a sum of smaller, simpler waves. This sum is called the Fourier series, and each simple wave equation is called a Fourier term (8).

Using any of the wave equations given above as Fourier terms, we can create a Fourier series of $n$ terms as follows:

$$f(x) = F_0 cos\ 2\pi(0x + \alpha_0) + F_1 cos\ 2\pi(1x + \alpha_1) + F_2 cos\ 2\pi(2x + \alpha_2) + F_3 cos\ 2\pi(3x + \alpha_3) + \ldots + F_n cos\ 2\pi(nx + \alpha_n)$$

or alternatively

$$f(x) = \Sigma^{n}_{h=0} F_h cos\ 2\pi(hx + \alpha_h).$$

Other mathematical operations, which will not be discussed in this paper, make it possible to simplify the Fourier series to

$$f(x) = \Sigma^{n}_{h=0} F_h\, e^{2\pi i\,(hx)}.$$

However, the Fourier series that is useful to crystallographers is expressed in a three-dimensional fashion, $(x,y,z)$, which is enough data for the construction of the electron density map of the target proteins (8). Therefore, the Fourier series can be written as

$$f(x,y,z) = \Sigma_h \Sigma_k \Sigma_l F_{hkl}\, e^{2\pi i\,(hx+ky+lz)}$$

which gives us the Fourier transform

$$f(x,y,z) = \int_h \int_k \int_l F(h,k,l)\, e^{-2\pi i\,(hx+ky+lz)}\, dh\, dk\, dl.$$

The Fourier transform helps us understand the way the scattered waves behave, and further derivations of it give us the data used in the construction of the electron-density map of the target protein (Fig. 13). This data consists of measurements of amplitude. Phase is defined as the timing of a wave's peaks and troughs in relation to that of the neighboring waves (2). The data concerning the phases of the waves tell us where there will be constructive or destructive interference.

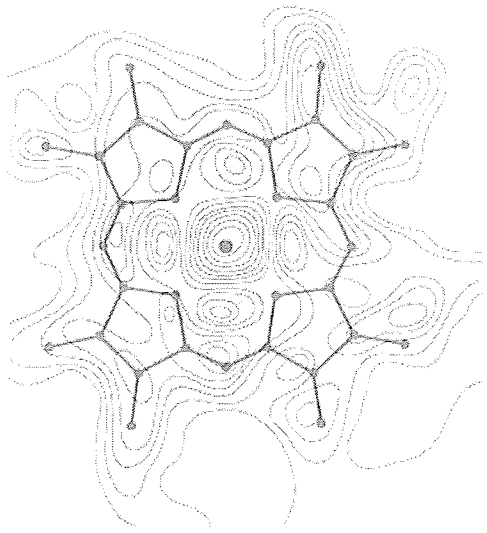

Fig. 13. Electron-density map of myoglobin.
From Stryer, p. 65.

**G. Conlcusion**

Considering that the mutation being discussed does not occur on the surface of the protein, the Y135F mutant should crystallize under the same conditions that yield crystals of the native retinol dehydratase protein. Screening of our protein with the Hampton Research Index solutions kit has yielded rosettes or spherulites (Fig. 8) under conditions that resemble those that yield crystals of the native retinol dehydratase protein. However, when trying to maximize these conditions, we have found no crystals. It is our belief that we are very close to getting this mutant to crystallize.

The Index solutions that yielded spherulites were solutions 84 and 85 which contained DHR Y135F incubated with PAP, EMTS and retinol. Solution 84 contained 0.2 M MgCl, 0.1 M HEPES pH 7.5, 25% w/v PEG 3350. Solution 85 contained 0.2 M MgCl, 0.1 M Tris Hydrochloride pH 8.5, 25% w/v PEG 3350. Not only

are these solutions similar to each other, but they are also very similar to those conditions that will crystallize the native DHR: 0.2 M $CaCl_2$, 0.1 M HEPES pH 6.6 – 8.1, 8 – 20% PEG 3350.

Further screening with the Hampton Research Wizard solutions kit may help us get closer to the conditions that will yield crystals. Future work with this mutant protein will also include the addition of mercury to the well solutions.

Once crystals can be obtained from the purified mutant protein, and data concerning the amplitude of each crystal unit cell generated by diffraction, the process of molecular replacement will probably be the one applied in the quest for the Y135F mutant's three-dimensional structure. This method consists of comparing the data obtained from my crystals with that of the already discovered structure of retinol dehydratase. This will allow me to have a good estimate of what the phases of my diffraction pattern may be, thus allowing me to generate an electron-density map.

**References:**

1. Stout, G. H.; Jensen, L. H. (1989) X-ray Structure Determination, A Practical Guide: Second Edition, Wiley-Interscience Publishing, Seattle, WA, 23.

2. Stryer, L. (1995) Biochemistry: Fifth Edition, W. H. Freeman and Company, New York, 63 – 65.

3. Branden, C.; Tooze, J. (1999) Introduction to protein structure: Second Edition, Garland Publishing, New York, 13, 375 – 376.

4. Pakhomova, S.; Luz, J. G.; Kobayashi, M.; Mellman, D.; Buck, J.; Newcomer, M. E. (2000) Acta Cryst., D56 1641 – 1643.

5. Pakhomova, S.; Kobayashi, M.; Buck, J.; Newcomer, M. E. (2001) Nature Struct. Biol., 8:447 – 451.

6. Grun, F.; Noy, N.; Hämmerling, U.; Buck, J. (1996) J. Biol. Chem., 271:16135 – 16138.

7. Vakiani, E.; Luz, J. G.; Buck, J. (1998) J. Biol. Chem. , 273;35381 – 35387.

8. Rhodes, G. (1992) Crystallography Made Crystal Clear, A Guide For Users Of Macromolecular Models, Academic Press, New York, 35—37.

9. Van Holde, K. E.; Johnsons, W. C.; Ho, P. S. (1998) Principles Of Physical Biochemistry, Prentice Hall, Upper Saddle River, NJ, 258 – 261.

*The quotes used concerning the Nobel Prize laureates were straight from www.nobel.se