12-19-2008

Application of X-ray Diffraction Methods and Molecular Mechanics Simulations to Structure Determination and Cotton Fiber Analysis

Zakhia Moore
University of New Orleans

Application of X-ray Diffraction Methods and Molecular Mechanics Simulations to Structure Determination and Cotton Fiber Analysis

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment
of the Requirements for the Degree of

Doctor of Philosophy
in
the Department of Chemistry

by
Zakhia Moore
B.S., Xavier University, 2000
M.S., University of New Orleans, 2005

December, 2008

ACKNOWLEDGEMENTS

Giving all glory to God for His many blessings, whom without I could not have completed this long journey. The work took longer than what I expected, but I'm glad it's finally behind me and now I can continue to another chapter in my life. This dissertation is dedicated to my beloved grandmothers, Dorothy Moore and Pernella Collins. Both of these beautiful women never stopped believing in me and always gave me words of encouragement. My parents, Vanest and Denise Moore have always been there for me every step of the way; I thank them for their constant unwavering love and support. To my best girls, Chanel Fortier, RaShante Gilford, Ebony Terrance, and Stephanie Anderson, there aren't words I could say that would come close to how much your friendship and love has helped me overcome many obstacles and brought me to where I am today. Thanks for everything all of you have given to me. To my co-workers, the U.S. Customs and Border Patrol Agriculture Specialists and Officers, too many to name, thanks for your support. My co-worker and editor, Dr. Yi He, thanks for your patience and for reading all this stuff! To Devron Thibodeaux, for showing me his passion for his work and family, thanks for being a great colleague and friend. Finally Dr. Edwin D. Stevens, you have put up with me for eight years, my ups and downs, the constant changes in my life. I thank you the most for not only being a great professor, but also a great advisor and friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE **PAGE**

ABSTRACT

The results of three very different studies are presented. X-ray diffraction has been utilized for single-crystal structure determinations, fiber diffraction analyses, and in conjunction with molecular modeling of Cellulose III_I. Although each technique is different in its sampling, data acquisition, data treatment, and identification, the common denominator has been the use of x-rays. The single-crystal structure determination of ethylene glycol bis(tropane-3-carboxylate) is presented as an example of the use of modern single-crystal x-ray instrumentation including the use of coupled charged devices (CCDs) as detectors for accurate data collection and rapid elucidation of crystal structures. The structure determination of Cellulose III_I by x-ray diffraction and computer modeling is presented to show how the use of x-rays in weakly diffracting materials can generate a reliable structure and be a key component in model building. Finally, a study is presented in which x-ray fiber diffraction data is utilized to investigate possible correlations between the crystallite orientation, crystallinity, crystallize size and the strength properties of cotton fibers collected from various countries.

Keywords: X-ray diffraction, structure determination, cotton fibers, molecular modeling

CHAPTER I

INTRODUCTION TO X-RAY DIFFRACTION

1.1 The Significance of X-ray Technology

X-ray technology has developed into one of the most notable methods of structural analysis during the past 100 years, as evidenced by how its discovery and enhancements play an important role in daily living. Versatility is an attribute of x-ray technology, offering widespread use in many applications. It is useful not only to the scientist, but also to health professionals and law enforcement officers. For example, x-rays are of enormous value in detecting and diagnosing health problems or assisting in pre-surgical procedures[1], as a convenient inspection device of luggage prior to boarding airplanes[2], and effective in detecting smuggled materials in cargo trucks or ships entering or passing through the nation's borders[3]. In particular, X-ray diffraction, based upon the scattering of x-rays, has become the premier technique for quantitative and qualitative analysis of crystalline materials, aiding in the new frontiers of nanotechnology and space exploration[4,5,6]. Most importantly, the determination of chemical structure of

various molecules is indispensable to chemists in an effort to gain insight into chemical problems[7].

Only a few physical methods are utilized to determine chemical structures, and amongst these methods, x-ray diffraction techniques have been the most successful. Diffraction methods yield atom positions, bond lengths, bond angles, and spatial proximity of non-bonded atoms for materials capable of forming crystalline solids.

The discovery of x-rays in 1895 by the German physicist Wilhelm Röntgen was quickly followed by the demonstration by von Laue[8] of diffraction of x-rays by crystals. With the addition of W.L. Bragg's diffraction theory in 1912[9], this non-destructive analytical technique has become extremely successful since it is one of few "fingerprinting" methods that can be used to accurately characterize both the identity and amount of compounds found in any crystalline system.

It is well known that any material which is made up of an ordered array of atoms will give a diffraction pattern. Determinations of the three-dimensional structure of compounds are most easily achieved by single-crystal x-ray diffraction. Single-crystal x-ray diffraction analysis differs from other diffraction methods because the measurement of the diffraction pattern is generated from an oriented single-crystal sample[10]. The diffraction pattern produced depends on the atoms present, their locations, and thermal motion. Modern experiments use an x-ray detector based on CCD camera technology, and the diffraction pattern from a single crystal yields a three dimension intensity distribution that appears as a series of "spots" in the detector image. Fourier series

analysis and least-squares refinement of the intensities of the spots allows accurate determination of the chemical identity and molecular structure of the sample.

X-ray powder diffraction (XRD) is an alternative analysis method that derives its name from the fact the specimen is in the form of a microcrystalline powder.

In XRD, the scattered signal contains the same information as the single-crystal experiment, but the three-dimensional pattern is “compressed” into one dimension. The diffraction pattern from a powder consists of “rings” of diffracted intensity with cone angles corresponding to the Bragg 2θ angles of each plane. Consequently, there is usually considerable overlap of peaks in the powder diffraction pattern, leading to severe ambiguities in extracting the intensities $I_{(hkl)}$ of individual diffraction maxima. As a result, XRD is rarely used for structure determination, except for inorganic compounds with relatively small cells and highly symmetric structures. On the other hand, XRD remains a very powerful technique for the identification and quantification of crystal phases with known structure.

X-ray fiber diffraction is a related technique used for structural analysis of fibrous materials (i.e. DNA, muscle fibers, cotton fibers, synthetic polymers) in which the ordering of the atoms is one-dimensional (along the fiber axis). The chain molecules in fibers are parallel to each other, but are usually randomly oriented perpendicular to the fiber axis, and usually terminate at random. Many fibers have only helical symmetry, in contrast to the three-dimensional symmetry seen in single crystals. In addition, depending on the conditions of crystallization and processing of the fiber samples, there may be varying degrees of misalignment of the fibers. As a result, a fiber diffraction

pattern typically consists of “arcs” of diffracted intensity, a pattern intermediate between single crystal “spots” and powder diffraction “rings”.

The investigations presented here demonstrate the application of various x-ray diffraction techniques to a variety of analytical and structural problems.

These studies demonstrate to ability of X-ray diffraction techniques to reach across many disciplines. A wide range of structural information can be obtained through X-ray diffraction, since it takes advantage of the scattering of x-rays by crystalline or partially crystalline materials[11]. In addition, when combined with other experimental or computational methods of analysis, it can provide deeper insight than either technique would provide on its own.

The results of three very different studies are presented. X-rays have been used for single-crystal structure determinations, fiber diffraction analyses, and in conjunction with molecular modeling of Cellulose III_I. Although each technique is different in its sampling, data acquisition, data treatment, and identification, the common denominator has been the use of x-rays. The single-crystal structure determination of ethylene glycol bis(tropane-3-carboxylate) is presented as an example of the use of modern single-crystal x-ray instrumentation including the use of coupled charged devices (CCDs) as detectors for accurate data collection and rapid elucidation of crystal structures. The structure determination of cellulose III_I by x-ray diffraction and computer modeling is presented to show how the use of x-rays in weakly diffracting materials can generate a reliable structure and be a key component in model building. Finally, a study is

presented in which x-ray fiber diffraction data is utilized to investigate possible correlations between crystallite orientation and the strength properties of cotton fibers collected from various countries.

1.2 Properties and Production of X-rays

Wilhelm Röntgen discovered a new form of radiation in 1895 and named it X-radiation to indicate its unknown character. X-radiation can pass through many materials that absorb visible light, including body tissues. X-rays also have the ability to knock electrons loose from atoms. They are characterized as a short-wavelength, high-energy form of electromagnetic radiation with a wavelength between 0.01 and 1nm and typical photon energies in the range of 100eV to 100keV[12]. Since their wavelength is comparable to the size of atoms, and they easily penetrate most materials, x-rays are ideally suited for investigating structural arrangements of atoms and molecules in a wide range of materials. Energetic X-rays can also penetrate deeply into materials and provide information about the bulk structure (x-ray radiography and tomography).

Generally, production of x-rays is achieved using sealed x-ray tubes, rotating anode systems, or synchrotron radiation. The primary source of x-rays in conventional laboratories is x-ray tubes, or “stationary” anodes. The traditional x-ray source consists of an evacuated glass bulb while more sophisticated tubes consist of a metal ceramic

envelope under vacuum. Inside the evacuated area, the tube contains a cathode consisting of a filament wire and an anode, which consists of a metal target with a high melting point. An electrical current drives electrons through the low resistance filament wire, which becomes hot and electrons are emitted. Due to a high voltage applied between the cathode and anode, emitted electrons are accelerated in the direction of the metal target.

On impact, electrons collide with atoms in the metal target and slow down, producing a continuous spectrum of x-rays, which is termed *Bremsstrahlung* radiation.

The electrons also eject inner shell electrons in atoms of the metal target through the ionization process. When an inner shell electron is removed, it is replaced by an electron from a higher-level shell. Consequently, radiation is released with a specific energy corresponding to the difference in energy levels between the initial and final states of the electron dropping into the lower energy shell. When a free electron fills the shell, an x-ray photon with energy characteristic of the target material is emitted. Thus, CuKα radiation arises when an electron in the L shell (n=2) drops in to the K shell (n=1). In this convention, Cu designates the target material, K designates the ground state electron shell of the transition and α designates $\Delta n=1$. Common targets used in x-ray tubes include Cu and Mo, which emits 8 keV and 14 keV x-rays with corresponding wavelengths of 1.54 Å and 0.7107 Å, respectively. The energy (E) of an x-ray photon and its wavelength λ are related by the equation $E = hc/\lambda$, where h is Planck's constant and c is the speed of light.

The rotating anode was invented in the 1960s in an effort to increase x-ray intensity and improve heat dissipation by spreading the electron bombardment over a much larger piece of metal[13]. The X-ray beam generated is more intense than those obtained from a stationary anode tube operated under similar conditions. The target metal is subjected to a focused stream of electrons originating from the cathode and accelerated by a high potential difference between the target disc and the cathode.

When the electron beam hits the anode, it produces an x-ray beam by the same mechanism as a sealed tube. However, only a very small portion of the energy of the electrons is converted to X-rays, the rest of the energy is converted into heat.

The anode rotates in vacuum and is internally cooled with water. The rotation continuously brings cooler metal into the path of the focused electron beam. A seal around the anode shaft maintains the vacuum while rotating and prevents leaks. Continuous pumping by a turbo molecular pump backed by a pre-vacuum pump maintains the high vacuum. With more efficient cooling, rotating anode systems can be run at a power almost an order of magnitude higher than systems equipped with an equivalent sealed tube.

Synchrotron radiation is inherently advantageous to laboratory sources since the naturally high-intensity, collimated beam provides superior resolution and easily tunable wavelengths. A synchrotron is a device that accelerates and steers electrons (or other elementary particles) by magnets in an evacuated ring[14]. Every accelerated charged particle produces some electromagnetic radiation. Synchrotron radiation is the name given to the electromagnetic radiation emitted by the charged particles circulating in a

ring. The diameter of the evacuated ring can be meters or miles in length. This occurs because the charged particles are accelerated (deflected) by the magnetic field from the dipole magnets to make the beam travel around the ring. A synchrotron produces a continuous distribution of infrared light, visible light, ultraviolet light and x-rays. Using single crystal monochromators, researchers are able to select the precise wavelength that they require from the continuous distribution of light produced.

Modern synchrotron radiation sources can generate highly energetic x-rays that are 10^{14} orders of magnitude brighter than the traditional x-ray source[15]. In most cases, synchrotron radiation is not practical for everyday chemical analysis, because of its huge size, cost, and location far away from local laboratories.

Synchrotron radiation technology is mostly used for special applications, including when a diffraction pattern needs to be achieved within minutes rather than hours per sample, or for fragile samples with little crystallinity where the best possible diffraction pattern is warranted.

1.3 Geometry of X-ray Diffraction and Bragg's Law

The x-ray diffraction pattern of a crystalline material serves as an identification tool and allows in some instances complete elucidation of its structure. Klug and Alexander simply described x-ray diffraction (where the interaction occurs between the electric vector of x-ray radiation and the electrons of the crystalline substance) as

billiard balls bouncing off one another[16]. X-rays are “scattered” by the electrons of the atoms without a change in wavelength. The electrons are believed to absorb and emit the impinging x-rays, i.e. the electron in the presence of electromagnetic waves will be excited to higher unstable energy levels. Upon relaxation, the electrons emit electromagnetic energy of the same frequency and wavelength. When x-ray photons collide with electrons, some photons from the incident beam will be deflected away from the direction where they originally traveled.

If the wavelength of these scattered x-rays did not change (meaning that x-ray photons did not lose any energy), the process is called elastic scattering (Thompson Scattering) meaning that only momentum has been transferred in the scattering process. In some directions, the scattered x-rays combine (crest to crest), which produces an increase in amplitude resulting in constructive interference and an increase in diffraction intensity. These are the x-rays that we measure in diffraction experiments, as the scattered x-rays carry information about the electron distribution in materials. In other directions, the out of phase combination of scattered x-rays results in destructive interference, and zero diffracted intensity. Also, in an inelastic scattering process (Compton Scattering), x-rays transfer some of their energy to the electrons and the scattered x-rays will have a different wavelength than the incident x-rays. These x-rays will contribute a slowly varying background radiation to the experiment.

For a given set of lattice planes with an inter-plane distance of *d*, the condition for a diffraction (peak) to occur can be simply written as:

$$n \lambda = 2d_{hkl} \sin \theta \qquad (1.1)$$

The English physicists Sir W.H. Bragg and his son Sir W.L. Bragg derived the equation in 1913 to explain why the cleavage faces of crystals appear to reflect x-ray beams at certain angles of incidence (theta, θ)[17]. They noticed the similarity of diffraction to ordinary reflection and treated diffraction as "reflection" from planes in the lattice.

In this equation, the variable *d* is the distance between atomic layers in a crystal, lambda (λ) is the wavelength of the incident x-ray beam, and n is an integer representing the order of the diffraction peak. In simple structures, the peaks in an x-ray diffraction pattern are directly related to the atomic distances through equation 1.1. Figure 1.1 shows an incident x-ray beam interacting with the atoms arranged in a periodic manner. The atoms, represented as spheres in the diagram below, can be viewed as forming different sets of planes in the crystal. Here, Bragg's Law illustrates that a set of parallel planes with index hkl and interplanar spacing d_{hkl} produces a diffracted beam when x-rays of wavelength λ impinge upon the planes at and angle θ and are reflected at the same angle.

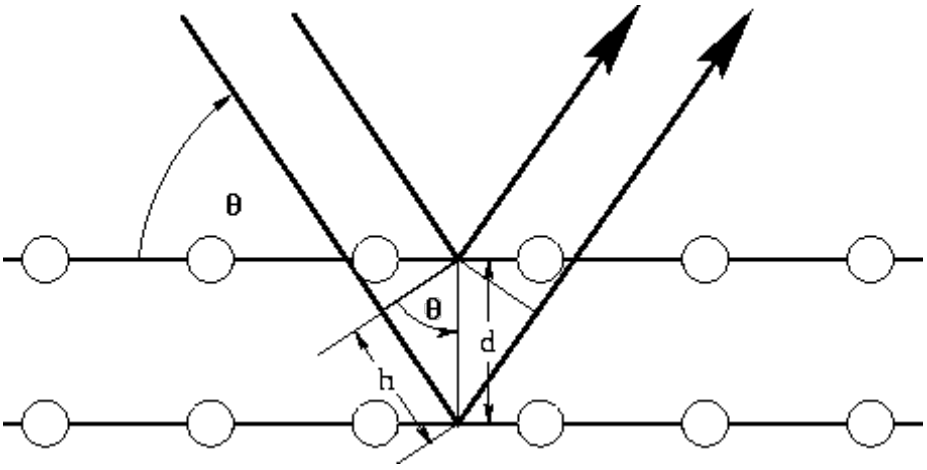


Figure 1.1. Bragg's Law

1.4 The X-ray Diffractometer

An X-ray diffractometer is a scientific instrument designed for the collection of accurate x-ray diffraction data. A diffractometer typically consists of an x-ray source, a goniometer for accurately mounting and orienting a sample, and a detector for collecting and counting scattered x-ray photons at a known scattering angle. For single-crystal experiments, the diffractometer usually includes a monochromator crystal to select a specific x-ray wavelength and a metal tube called a collimator that directs a narrow incident x-ray beam onto the sample. To orient the single-crystal sample, a goniometer allows rotation of the sample about 2 or 3 independent axes under computer control. The detector is also mounted on an axis that allows variation of the total scattering angle (2θ) under computer control.

The instrument often includes a low-temperature system for cooling the crystal sample which reduces thermal motion of the atoms and improves the resolution of the experiment.

A diffractometer designed for powder diffraction measurements is less complex because the randomly oriented microcrystalline nature of the sample eliminates the need to place the sample in a particular orientation with respect to the incident beam. The diffractometer generally allows for rotation of the sample (θ) and detector (2θ) about a common axis. For fiber diffraction measurements, the diffraction pattern is 2-dimensional and the scattering intensity much weaker than single-crystal or powder diffraction intensities. Efficient measurement requires the use of a rotating anode or synchrotron x-ray source and a two-dimensional area detector.

1.5 Area Detectors

Many important problems in solid physics, biophysics, and materials science can be studied by means of x-ray diffraction[18]. In the past, important contributions to the understanding of these problems have resulted from the application of recording techniques utilizing photographic emulsions and single point electronic detectors[19]. As attention had been directed to more difficult problems in these areas, both methods have become to be decreasingly practical.

Important new problems often exhibit one or more of the following characteristics:

1. The sample is weakly diffracting; thus a high efficiency detector is required.
2. The pattern consists of many diffracted beams in two dimensions; therefore

an area detector is required.

3. The samples changes with time, due either to its dynamic characteristics, or to the effects of the radiation it is receiving.
4. It is necessary that the detector is capable of recording high count rates, i.e. it must not be “count-rate limited.

New techniques have been developed and designed to meet the uniqueness of these more difficult situations. Given the above information, X-ray scattering studies of large, weakly diffracting materials require detectors that have good spatial resolution and very low noise levels. It is also very efficient to collect a full two-dimensional diffraction pattern all at once-with an “area detector” rather than simply measuring the intensities at one point or along one line at a time. X-ray film is commonly used as an area detector for CuKα (8 keV) x-rays[20], but it has several significant limitations: it has a high background noise level which leads to a very poor detective quantum efficiency for weak signals; it must be developed and then digitized before the data may be analyzed quantitatively; and it has a very limited dynamic range, so that a typical pattern must be recorded on several films that are exposed for different times and then scaled together.

Whereas traditional diffractometers use point detectors, which measure the intensity of each diffracted beam individually in sequence, area detectors record the diffraction pattern over a large area of reciprocal space simultaneously. One type of area detector is the multiwire or gas proportional detector. An example of this type is the Hi-Star detector located in the UNO Chemistry Department which consists of two

perpendicular sets of parallel wires in a flat box filled with an xenon gas (see Figure 1.2). A thin window of beryllium permits entry of x-ray from the front of the detector[21].

Entering the detector through the beryllium window, an x-ray photon ionizes the gas in a small region, producing a few hundred electrons. The electrons drift to the nearest anode wire, and because of the high voltage, each electron triggers an electrical discharge that in turn produces thousands of ion pairs in the gas. The movement of these ions in the electric field of the cathode and anode wires produces a pulse of current in each of the nearest wires. The detection of these pulses at the ends of the x and y delay lines allows determination of the reflection position in the detector. The output from the area detector is fed to a computer, which indexes the event using the x and y positional information and the crystal orientation at the time of the event. The computer sums events that have the same index and thus produce a file of indexed intensities.

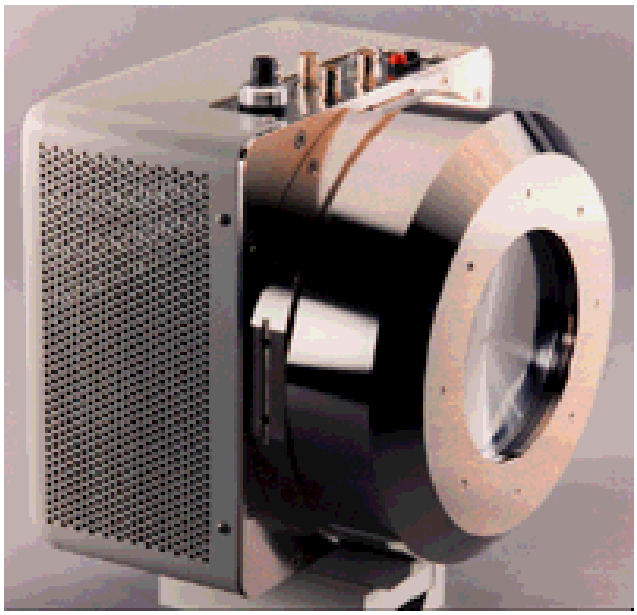


Figure1.2. Bruker AXS HI STAR Area Detector.

Another type of area detector utilizes charge-coupled devices (CCDs). A CCD area detector consists of a collection of individual photoelectric sensitive elements, called pixels, that are arranged linearly or in a two-dimensional pattern on a single semiconductor chip[22]. The chip, which is usually silicon and typically has dimensions of a few millimeters on a side, also contains electronic circuitry that makes it possible to determine the electrical output signal from each of the photosensitive elements either sequentially or simultaneously. To record an x-ray image using a CCD based area detector, the x-ray image is usually first converted to a visible light image using a phosphor screen. The phosphor's function is to convert x-ray energy into visible light while preserving the spatial content of the x-ray image. The visible light image formed by the phosphor is focused by a lens, or transferred by the fiber optic taper, onto a CCD chip to generate an electronic image, corresponding to the original x-ray image, which then can be digitized, saved, analyzed, and displayed. The processes in this energy conversion can be described as:

1. Absorption of an x-ray photon by the phosphor and formation of an excited state,
2. Partial relaxation of the excited state by radiationless decay,
3. Radiation by emission of a lower energy photon from the excited state, and
4. Relaxation to the original, ground state.

There are several reasons for using a phosphor rather than having x-rays strike directly on the CCD.

First, silicon only weakly absorbs x-rays with energy higher than 5 keV. These x-rays are inefficient in directly forming an electronic image, since most of them simply pass through the active region of the CCD. This problem can be avoided by using a phosphor screen of heavier atoms, which strongly absorbs the x-rays and efficiently converts them to visible light. Second, for the x-ray photons that are absorbed by silicon atoms in the CCD, each one will generate thousands of signal electrons. The dynamic range can be improved if an energy converter is used with the CCD, e.g., a phosphor screen and fiber optic taper which produce on the order of ten signal electrons in the CCD for each x-ray photon absorbed in the phosphor. Third, a phosphor screen and its optically transparent substrate will stop most of the x-rays and will help to protect the optical system and the CCD from radiation damage. Finally, a phosphor screen coupled to a CCD with a demagnifying lens or fiber optic taper can record images much larger than the active area of the CCD.

In the UNO Chemistry Department, a Bruker AXS SMART CCD diffractometer system is equipped with a SMART 1000 CCD detector, which has a pixel size of 120 microns containing 512 X 512 pixels. For data collection, CCDs are coated with phosphors that emit visible light in response to x-rays. A tapered bundle of optical fibers are used to increase light collection efficiency between the phosphor and the CCD. At the end of each collection cycle, the charges are read out by a process in which rows of pixel charge are transferred sequentially into a serial readout row at one edge of the CCD.

After the charges in the readout row are transferred serially to an amplifier at the end of the row, and the next row of pixel charges will be transferred into the readout row. Because all data are read out at the end of the data collection, a CCD has no dead time, and thus no practical limit on its rate of photon counting.

CHAPTER II

IDENTIFICATION AND STRUCTURE DETERMINATION OF ETHYLENE GLYCOL BIS(TROPANE-3-CARBOXYLATE) BY SINGLE-CRYSTAL X-RAY DIFFRACTION

2.1 The Use of Single Crystal X-ray Diffraction in Structure Determinations

"Structure determines function" is an axiom which emphasizes the connection between how a molecular system functions and its natural structure. The main objective of x-ray crystallography is to obtain knowledge of the molecular structures of natural and synthesized compounds. It yields the three-dimensional structures of new and existing crystalline materials, which can be used to predict or interpret functional information. More specifically, x-ray crystallography reveals what atoms are present and their positions, distances and angles between atoms, and the symmetry involved that generates the entire crystalline substance. Single-crystal structure determination has become an important and extremely powerful tool, not only for mineralogists, inorganic and structural chemists, but also for many other scientists who are interested in the structural basis for the properties of chemical and biological systems at the molecular level.

Single-crystal x-ray structure determination may also be regarded as the ultimate analytical tool, because it provides direct, unequivocal identification of the sample under investigation. Unlike many spectroscopic techniques, a successful x-ray structure

determination yields the precise composition of the unit cell, including the identity and position of every atom.

Except in rare cases, only the correct molecular structure will produce a good fit to the observed x-ray diffraction intensities, and all incorrect structures produce fits that are obviously inferior.

A number of other experimental techniques may also provide useful structural information, but they are often limited in the amount of information or resolution they can provide, or suffer from other limitations. Electron diffraction and microwave spectroscopy can provide very accurate structural data for molecules in the gas phase. Other forms of spectroscopy, including Extended X-ray Absorption Fine Structure (EXAFS) and Resonance Raman spectroscopy provide only limited information about the environment of a small number of atoms, bonds, or functional groups, mostly of elements having atomic numbers greater than oxygen[23].

Libraries of spectra are helpful in identifying compounds and commonly known functional groups. However, this practice may not be useful in the investigation of newly synthesized compounds not found in spectral libraries. Techniques for the direct visualization of molecules, such as atomic force microscopy, in which a probe is scanned over molecular surface, do not provide details of the molecular interior.

Over the past few decades, many technological improvements have been directed to optimizing X-ray diffraction instruments and software programs, so that the once long, arduous task of structure determination has evolved into a fairly straightforward analytical technique. Single-crystal analysis differs from other diffraction

methods because the measurement of a diffraction pattern is determined in three dimensions, and generated from an oriented single crystal[24].

It is the only diffraction technique where the two-way mathematical relationship that exists between the observed diffraction pattern and the structure of the scatterer, which is the electron density distribution of the crystal, is routinely achieved in practice. The key assumption is that a single crystal has three-dimensional translational symmetry, which reduces the problem to a study of the scattering density of a unit cell, rather than the entire crystal.

Nuclear Magnetic Resonance (NMR) spectroscopy is the technique most competitive with X-ray crystallography for generating three-dimensional structure information of macromolecules. With NMR, one can obtain distances between specific nuclei in the structure. When combined with molecular dynamical or molecular mechanical techniques, these data can be used to produce full three-dimensional molecular models. However, even with the best available equipment, the size of structures that can be solved by NMR is limited. The structures of viruses, complex proteins and enzymes that are routinely being solved by X-ray diffraction are currently beyond the capabilities of NMR methods. An advantage of NMR, however, is that experiments can be carried out on samples in solution, avoiding the need to grow crystals.

Unfortunately, X-ray crystallography is not appropriate in every situation. For example, some molecules of interest may fail to crystallize, while others even though

crystallized will not diffract well. Moreover, the phase problem (discussed later) can be challenging. Special circumstances like twinning (not discussed) can also interfere with data collection and analysis. However, crystallography is a complex but valuable technique, which certainly requires specialized skills, experience, and patience.

The successful structure determination of ethylene glycol bis-(tropane-3-carboxylate), a ditropane derivative synthesized as a potential ligand for neuronal nicotinic acetylcholine receptors (nAChRs) for treating various central nervous system diseases, is presented in full detail[25]. We will describe every step involved in this particular structure determination, from sample selection to model building and refinement. The crystal was a kind gift from Dr. Mark Trudell's research group, Chemistry Department, University of New Orleans.

2.2 Description of Unit Cells and Lattices

Crystals are solid-state materials in which the atoms or molecules have a repeating order which extends over a long range[26]. The repeating order in a crystal can be simply described as if they were the two dimensional patterns printed on a piece of wallpaper. Most wallpaper has a regular repeating design that extends from one end to the other. Crystals have a similar repeating design, but in this case the design extends in three dimensions from one edge of the crystal to the other. We can easily describe a piece of wallpaper by specifying the size, shape, and contents of the simplest repeating unit in the design. Similarly, we can describe a three-dimensional crystal by specifying the size, shape, and contents of the simplest repeating unit and the way these repeating

units stack to form the crystal. The repeating unit in a crystal is called a unit cell. The unit cell is regarded as the basic "building block" of a crystal. It is a human construct used to simplify our notion about the size, dimensions, and number of objects contained within the crystal.

Theoretically, we should be able to reconstruct the entire crystal just by placing a large number of these unit cells next to each other in all directions.

Each unit cell is defined in terms of lattice points. The lattice is the basic network of points on which the repeating unit (the contents of the unit cell) may be imagined to be laid down so that the regularly repeating structure of the crystal is obtained. Thus the lattice establishes the repeating pattern, and the unit cell tells us what is being repeated. In Figure 2.1, we see a standard three-dimensional unit cell consisting of a parallelepiped with cell edges of length **a**, **b**, and **c**, and angles of **α**, **β**, **γ**. The unit cell is always chosen to be right-handed, with a, b, and c following the “right-hand rule,” the direction of the cell edges will be chosen to coincide with the major symmetry elements within the unit cell. All crystal structures fall into one of the seven crystal systems, its restrictions are according to its unit cell lengths and angles.

Table 2.1. Laue Lattices

Cell Type	Axial Measurements	Angle Measurements

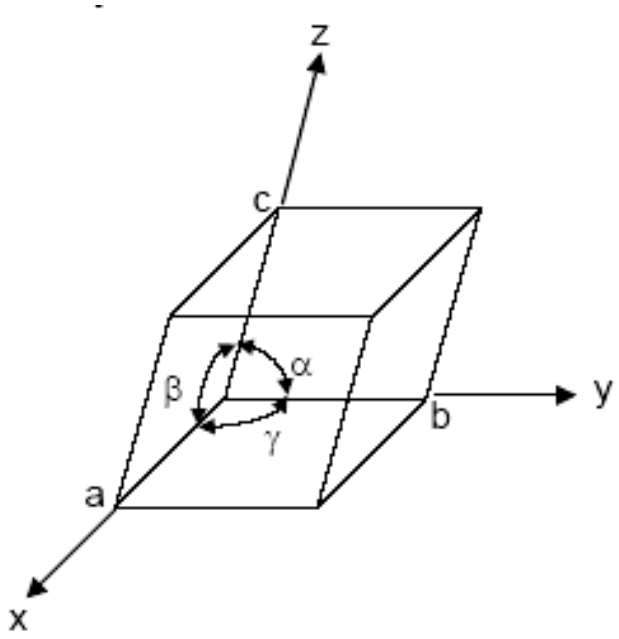


Figure 2.1. A three-dimensional unit cell.

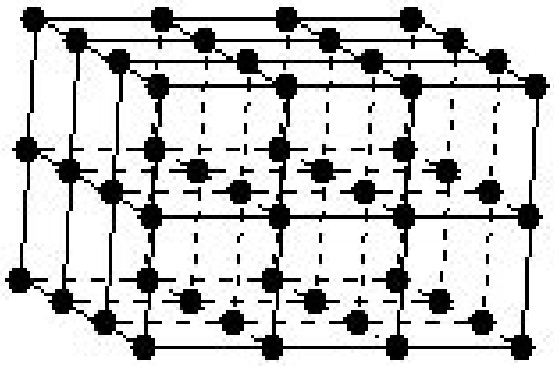


Figure 2.2. Three-dimensional lattice

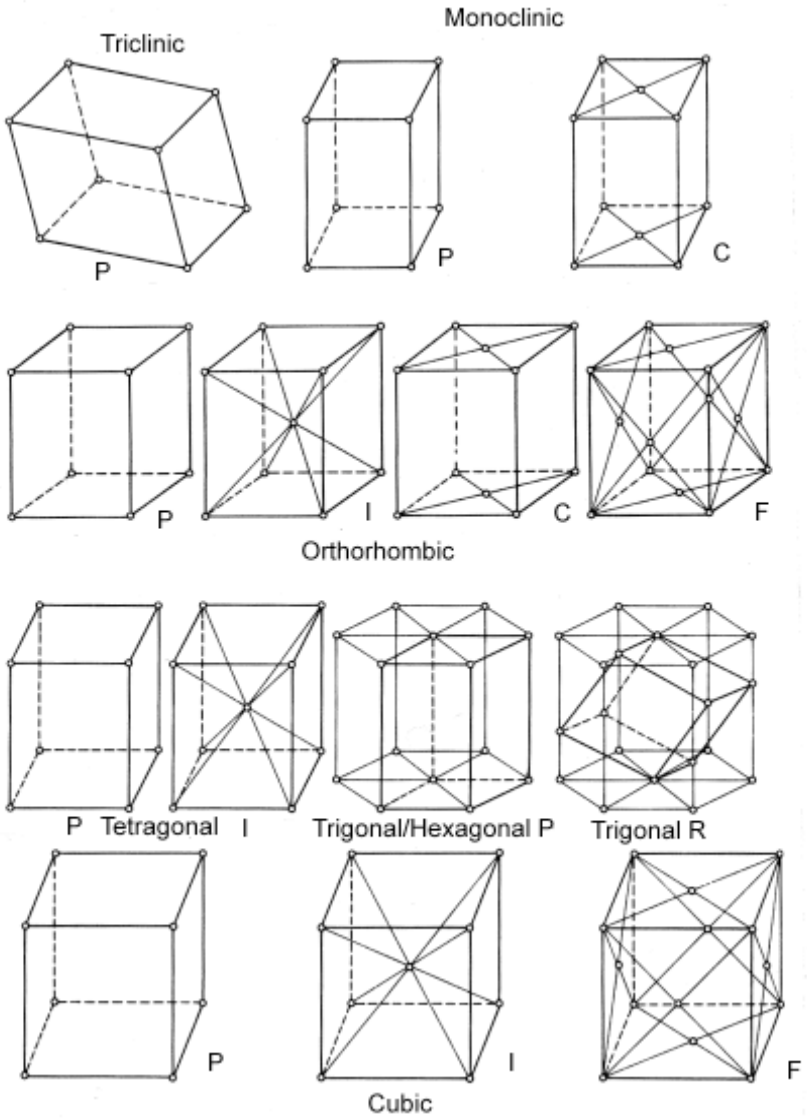


Figure 2.3. Bravais Lattices.

2.3 Miller Indices

What are Miller Indices, and how are they important in determining crystal structure? Bragg's law relates the angle of x-ray diffraction to the interplanar spacing d_{hkl} of a specific set of repeating planes in the crystal structure. Miller indices (h, k, l) specify the orientation of the crystal planes by the intersection of the plane with the crystallographic axes (**a**, **b**, **c**) of the solid. Each plane has integer indices *h*, *k*, and *l* equal to the reciprocals of the intercepts *a*, *b*, and *c* as fractions of the unit cell dimensions. Miller indices are also used to identify a reflection (diffracted intensity I_{hkl}) coming from a set of hkl planes in a crystal. In Figure 2.4, the plane divides a, b, and c axis into whole units, so the h, k, and l index is 1. The Miller indices for this plane are (1 1 1). In Figure 2.5, we see that the plane parallel to the unit cell axes a and b which intersects the c axis at one half the unit cell length is assigned the Miller indices (0 0 2).

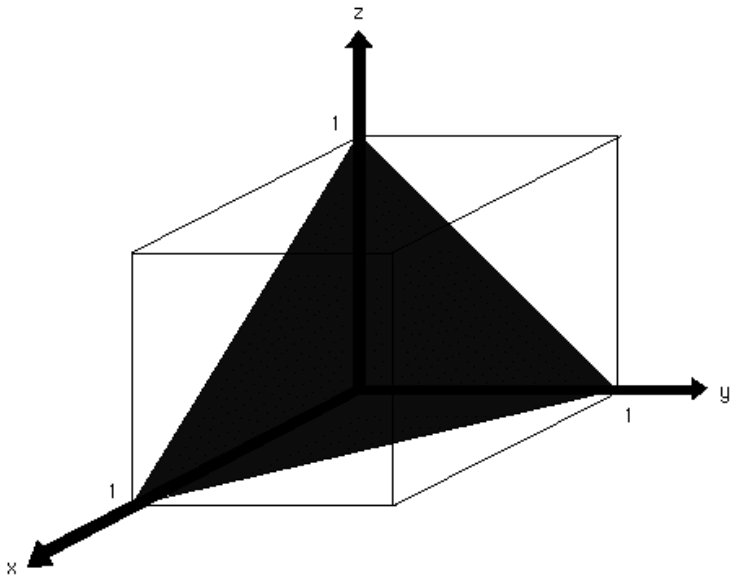


Figure 2.4. Miller Indices (1, 1, 1)

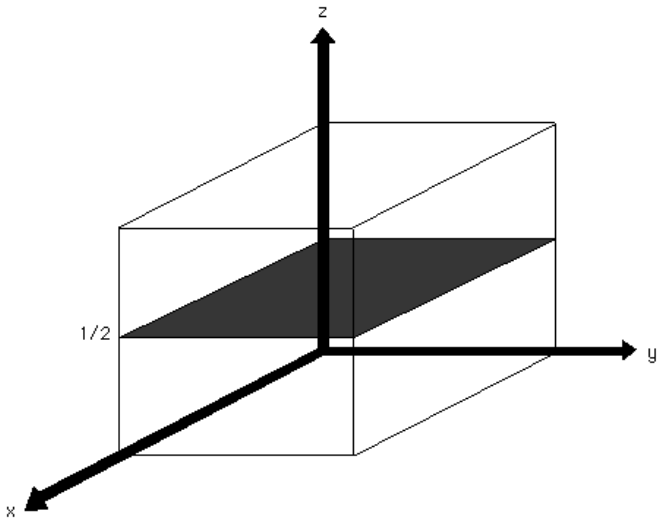


Figure 2.5. Miller indices (0, 0, 2)

2.4 The Reciprocal Lattice and Ewald Sphere

A single diffraction event (reflection) occurs when an entire set of parallel planes constructively interfere to produce the diffracted x-ray beam. The use of constructs such as the reciprocal lattice and Ewald sphere aid in determining geometrically where the reflections will occur and satisfy Bragg's equation. The reciprocal lattice is related to the real crystal lattice (axes **a**, **b**, **c**) with axes **a***, **b***, **c***, such that **a*** (**a*=**1/**a**) is perpendicular to **b** and **c**, **b*** (**b*=**1/**b**) is perpendicular to **a** and **c**, and **c*** (**c*=**1/**c**) is perpendicular to **a** and **b**. Each reciprocal lattice point corresponds to a set of Miller indices, hkl. Considering Bragg's equation, the angle of diffraction **θ** is inversely related to the interplanar spacing d_{hkl}. This means that large unit cells will produce small angles of diffraction, resulting in many reflections at a convenient angle from the incident beam. The opposite is true for small unit cells, which will produce fewer reflections. P.P. Ewald developed a geometrical construction to help visualize which Bragg planes are in the correct orientation to diffract. In reference to Figure 2.6, the Ewald sphere has a radius equal to $1/\lambda$, with its center *C*. Points *P*, *B*, and *O* are on the sphere. As the crystal is rotated about point *O*, a reciprocal lattice point *P* comes in contact with the circle. As incident x-rays passes through the crystal (line XO´) at an angle θ, the reflected x-ray diverges from point *C* at angle 2θ through point P_{101}. The lines *OP* and *BP* are drawn; the length of *OP* (or 000 to 101) is $1/d_{hkl}$. The length of *OB* is $2/\lambda$, the diameter of the sphere. The angle *BPO* is equal to $\sin\theta$.

$$\sin \theta = OP / BO = (1/d_{hkl})/ (2/\lambda) \quad (2.1)$$

Rearranged, this equation gives Bragg's law. The Ewald Sphere and the reciprocal lattice show that when a reciprocal lattice point falls on the sphere, a reflection will occur, thus Bragg's law is satisfied. Ewald's sphere shows which *hkl* planes are in the proper orientation to diffract, and how each reciprocal lattice point must be arranged with respect to the x-ray beam.

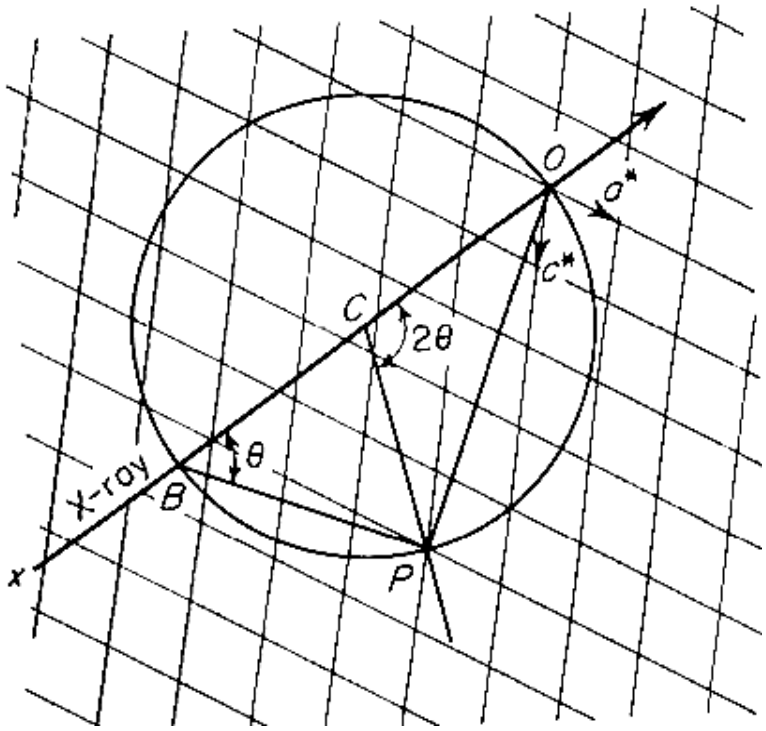


Figure 2.6. Diffraction in terms of the reciprocal lattice.

2.5 Crystal Selection

A suitable crystal is necessary for structure determination, as crystal quality is directly proportional to the quality of the diffraction pattern. A quality crystal should be appropriate in size and shape, without defects or obvious twining. Smooth faces and straight edges are useful guides in choosing a crystal for structure determination[27]. Since X-ray absorption reduces the intensities of the spots and introduces systematic errors into the intensity measurements, there is a limit to the size of the crystal. However, the absorption also depends on the x-ray wavelength chosen. During crystallization, randomly arranged ions, atoms or molecules originally in the gas phase or solution adopt a unique position, orientation, and symmetry in the unit cell. The resulting crystal is an orderly three-dimensional array of molecules, often held together by noncovalent interactions. The crystals grown for this particular structure determination were obtained by a slow, controlled recrystallization from methanol solution that evaporated over time[28]. Another way to determine crystal suitability is to place the crystal under a polarizing microscope to judge optical clarity. When the selected crystal was rotated, while being observed with plane-polarized light, the crystal rapidly changed from uniformly bright to dark and back again every 90^{o}. This behavior indicated the crystal likely consisted of a single domain with a common orientation.

The size of the crystal chosen for structure determination was approximately 0.26mm x 0.08mm x 0.03mm, which is appropriate for the 0.5mm x 0.5mm dimensions of the uniform portion of the x-ray beam.

This allowed all parts of the crystal to be illuminated in the beam with equal intensity. A linear absorption coefficient of 0.086 mm^{-1} was calculated using the equation:

$$I = I_o e^{-\tau\rho(\mu/\rho)\lambda} \qquad (2.2)$$

where I_o is the incident's beam intensity and $(\mu/\rho)\lambda$ is the mass absorption coefficient for the 0.71073 Å Mo wavelength used and ρ is the density. The density was calculated using the atomic molecular weight of the crystal and the volume of the unit cell. Since we are dealing with atoms of small atomic number, the need to correct for absorption can be ignored. Generally speaking, the absorption of x-rays from a crystal of such small thickness and small absorption coefficient should generate some systematic errors in intensity, but such errors are estimated to be well below the random noise level of the intensity measurements.

The preparation of the crystal for structure determination entailed placing the crystal at the end of a thin glass fiber attached by silicon vacuum grease. The fiber was fixed onto a brass pin and this pin was then placed on a goniometer head as shown in Figure 2.9. The goniometer head is a highly accurate xyz-positioning device for centering the crystal on the goniometer[29]. The crystal can be rotated about its mounting axis and another axis perpendicular to it. Translation adjustments along the x, y, or z-

axis allow the crystal to be centered and rotated through 360^{o}. These heads are easy to adjust and hold the crystal stationary during data collection. It is of the utmost importance that the crystal does not move during data collection. The crystal was centered on a Bruker SMART 1K x-ray diffractometer, and the orientation and unit cell dimensions of the crystal were determined by gathering a small set of preliminary data.

Equally important is the cooling of the crystal, as the atoms within the crystal are not at rest, but are constantly vibrating about their rest positions due to thermal motion. The greater the temperature of the crystal leads to larger amplitudes of atomic vibration, and consequently weaker intensities of the reflections. As a consequence, atomic positions and other structural results will be less accurate if the crystal is not cooled. Cooling is achieved by directing a stream of cold nitrogen gas down on the crystal. The nitrogen stream is generated by boiling liquid nitrogen, and the temperature monitored by a copper-constantan thermocouple mounted approximately 1.0 cm upstream from the crystal.

2.6 Diffraction of X-rays by Crystals

The diffraction of x-rays by a single crystal leads to a set of intensity data that can be used to determine the spatial arrangement of the atoms that make up the crystal. Diffraction is a suitable technique for x-rays because of the limitations of focusing optics of the relevant wavelengths. In order for the object to diffract light and thus be visible under magnification, the wavelength (λ) of the light must not be

significantly larger than the object. Visible light, which is electromagnetic radiation with wavelengths of 400-700 nm, cannot produce an image of individual atoms in molecules, in which bonded atoms are only about 0.15 nm or 1.5 Å apart. Electromagnetic radiation of this wavelength falls into the X-ray range, so even the smallest molecules diffract X rays. Even though individual atoms diffract x-rays, it is still not possible to produce a focused image of a molecule, since existing lenses cannot focus X-rays. However, by measuring the directions and intensities of the diffracted x-rays, a computer can be used to simulate the effects of an objective lens by calculation.

To determine the position of atoms from crystallographic data, the computer simulates the action of a lens, computing the electron density within the unit cell from the list of Miller indexed intensities. The Fourier transform describes precisely the mathematical relationship between an object and its diffraction pattern, which allows us to convert the distribution of reflections intensities into a Fourier-series description of the electron density distribution of the crystal. The intensity of an x-ray reflection can be described by the structure-factor equation, containing one term for each atom (or each volume element) in the unit cell. In turn, the electron density is described by a Fourier series in which each term is a structure factor. The crystallographer uses the Fourier transform to convert the structure factors into the electron density distribution, $\rho(x,y,z)$.

When incident x-rays strike a crystal, the electrons of each atom will absorb and immediately reflect the x-rays, radiating in all directions. The reflections are treated similar to simple waves, in which each function will have a different phase since the scattering is coming from different positions in the unit cell:

$$f(t) = f \cos 2\pi (\Phi t + \alpha) \qquad (2.3)$$

$$\text{or} \quad f(t) = f \sin 2\pi (\Phi t + \alpha) \qquad (2.4)$$

where f is the amplitude of the wave, Φ is the frequency, and α is the phase. A phase is the difference in position of the crests of two waves of the same wavelength traveling in the same direction. In the x-ray experiment, the intensity I = f*f of the scattered wave can be measured during data collection; however the phase information is lost.

The x-ray structure factor F_{hkl} represents the amplitude of the diffracted x-rays resulting from the sum of the scattering from all of the atoms in the direction defined by the Miller indices, *hkl*. Reflection F_{hkl} is calculated:

$$F_{hkl} = \sum_{j=1}^{atoms} f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \qquad (2.5)$$

or

$$F = |F| e^{i\alpha} \qquad (2.6)$$

As mentioned, the structure factor that describes reflection *hkl* is a Fourier series of atomic structure factors in which each term is the contribution of each atom in the unit cell, with is own amplitude, f_j, whose frequency is h in the x-direction, k in the y-direction, and l in the z-direction. For each possible set of values h, k, and l, the associated wave has amplitude F_{hkl} and phase α_{hkl}. The exponential term having both sine and cosine components describes the phase and frequency.

Since a single crystal has three-dimensional translational symmetry, it reduces the problem to a study of the electron density of the unit cell, rather than the entire crystal. The relationship that exists between the structure factor and the atomic coordinates is that the structure factor is the Fourier transform of the electron density:

$$F_{hkl} = \int_V \rho(x,y,z) \exp[2\pi i(hx + ky + lz)]\, dV \qquad (2.7)$$

where V is the unit-cell volume. The inverse Fourier transform is

$$\rho(xyz) = (1/V) \Sigma F_{hkl} \exp [-2\pi i (hx + ky + lz)] \qquad (2.7)$$

which means that the electron density (ρ) at any point (x,y,z) in the unit cell can be computed by summing over all structure factors (F), measured at the diffraction points identified by the integers *h,k,l*. By calculating the electron density of the unit cell, the atom positons can be found.

The x-ray structure factor F_{hkl}, is a complex number and can be expressed in terms of its amplitude and phase,

$$F_{hkl} = |F_{hkl}| \exp(2\pi i \alpha_{hkl}) \qquad (2.8).$$

Although we know how to calculate the electron density from F_{hkl}, only the indices of each reflection and its intensity are measured. The phase of F_{hkl} is lost during data collection. The phase is needed to calculate the electron density, hence the notorious "phase problem" of x-ray crystallography is created.

There are several techniques to estimate approximate phases, and from them calculate initial electron density maps. One of the first methods developed was the heavy atom method. This method requires the presence of a single or small number of heavy atoms whose positions can be determined using of Fourier series summation with coefficients $(F_{hkl})^2$ (the Patterson function). In recent years, the heavy atom method has been largely replaced by "Direct Method" techniques to solve the phase problem. Direct methods techniques rely on statistical relationships between the x-ray structure factors magnitudes that exist because the electron density is a function that must be everywhere equal to or greater than zero[31]. Some relief is given to the phase problem if the crystal structure is centrosymmetric (i.e., for every point (x, y, z) in the unit cell there is an indistinguishable point (-x, -y, -z)), then the phase is either positive or negative ($\alpha = 0^o$ or 180^o). The phases are determined by statistical relationships between certain reflections, the highest in amplitude having the most weight.

In multiple solution methods, the phases of a small number of normalized structure factors, E_{hkl}, are arbitrarily assigned positive or negative values. According to the principle of positivity, the signs of three reflections are related by:

$$s(h_1, k_1, l_1)s(h_2, k_2, l_2) \approx s(h_1 + h_2, + k_1 + k_2, l_1 + l_2) \qquad (2.9)$$

where the three reflections are chosen such that the indices of the third are the sums of h's, k's, and l's of the first and second reflections. The phase of the third reflection is equal to the product of the phases of the other two reflections, with a probability that can be calculated. All possible combinations of the starting set phases are used, and as many additional phases as possible are calculated. The E values are then used to

generate a set of E-maps, which are normalized electron density maps of estimated atomic positions for each set of possible initial phases. Usually, one of the phase sets will be more self-consistent that the others, and the resulting E map usually contains the correct molecular structure. If not, another phase set may be chosen. After initial positions of the atoms are located in the E map, better estimates of the atomic positions can be obtained by least-squares refinement. In this study, since the molecule did not contain any heavy atoms, direct methods were used to determine the initial phases, and all non-hydrogen atoms were located in a subsequent E map.

2.7 The Crystallography Project

The major steps in determining the crystal structure of ethylene glycol bis(tropane-3-carboxylate) included:

1. Unit cell determination
2. Data Collection
3. Integration of intensity
4. Generation of a trial structure
5. Refinement of the Crystal Structure

a. Unit cell determination

Knowing the unit cell dimensions prior to data collection is essential to devising a strategy that will give us as many identifiable (by index) measurable reflections as possible. It also indicates whether or not the crystal is actually suitable for further data collection. For example, if the crystal is a weak scatterer of x-rays, should the crystal be discarded, or will the study be successful if longer than usual exposure times are used. Once a crystal is chosen and mounted for data collection, a suite of computer programs is available to perform the formerly arduous task of structure determination. SMART is a part of a suite of programs included with the Bruker AXS SMART CCD-based X-ray crystallographic system[30]. This online program controls the x-ray diffractometer in order to collect the diffraction data used by the other programs in the system. SMART controls the x-ray shutter, and crystal orientation setting angles 2-Theta, Omega, and Phi (Figure 2.7), and storage and readout of the CCD detector.

After verifying detector calibration files, the crystal is mounted and 512 X 512 pixel images (frames) are collected (Figure 2.8). A frame is a 0.3° scan about the omega axis; for preliminary scans to determine unit cell values, a ten second frame is suitable. Typically, 3 scans of 25 frames are collected at different phi and theta angles to sample reflections in different regions of reciprocal space. During the scan, a low temperature system directs a stream of cold nitrogen gas over the crystal. This decreases the amplitude of atomic vibrations in the crystal, which increases the intensity of x-ray scattering, especially at high scattering angles. After the initial scan, the SMART program searches the frame data for intensity maxima (reflections) and determines the precise angles at which the scattering occurs. This information is used to determine the translational symmetry of the reciprocal lattice, which determines the orientation and dimensions of the crystallographic unit cell. Integer values of h, k, and l are then assigned to each reflection (indexing). Least-squares refinement of the observed setting angles yields unit cell dimensions (with estimated standard deviations) and crystal orientation parameters which give the best fit to the x-ray observations. Based on the unit cell dimensions, a tentative selection of the Bravais lattice type is usually made. The unit cell parameters and orientation matrix are written to a .p4p file for subsequent use by the other structure determination programs. Part of the .p4p output file for ethylene glycol bis(tropane-3-carboxylate) is given below in tabular form.

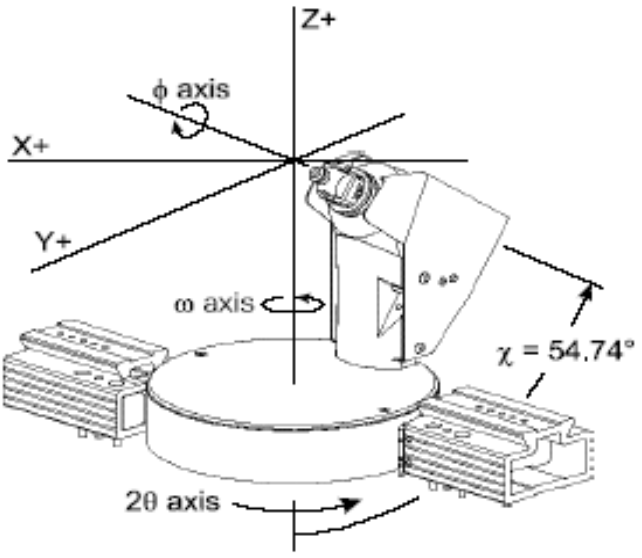


Figure 2.7. Goniometer axis showing all swing angles.

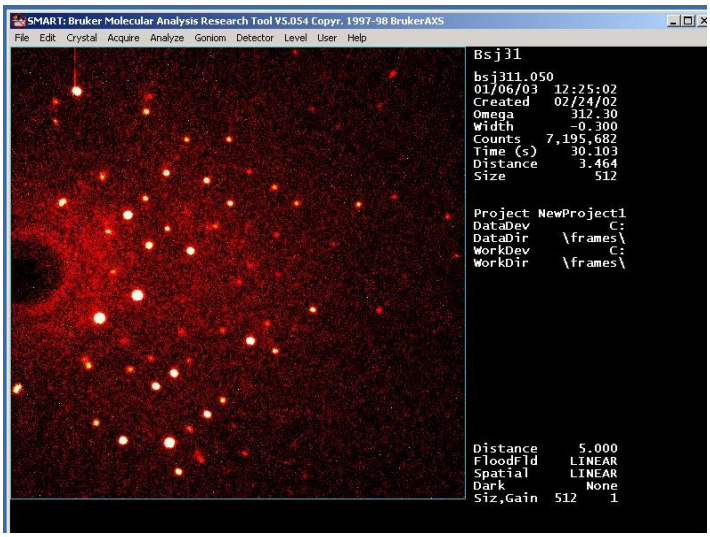


Figure 2.8. 512 X 512 pixel image. Image obtained from the actual data collected from crystal

Table 2.2. Summary of Unit Cell Parameters and Observed Reflections

Radiation Type	Molybdenum λ = 0.71073 Å
Source Power	50.00 kV, 16.00 mA
Crystal Color	Colorless
Crystal Size	0.26 mm 0.08mm 0.03mm
Data Collection Temperature	-123° C
X, Y Beam Center	247.2860 261.4540
Crystal Lattice	Monoclinic (b-unique) Primitive

Observed Reflections:

H	K	L	2-Theta	Omega	Phi	Chi	Intensity	I/sig
6	3	1	332.00	327.398	0	54.799	2372.3	42.0
1	-3	-10	332.00	329.165	0	54.799	76950	244.0
-1	1	0	332.00	328.595	0	54.799	89442	264.0
5	1	8	332.00	329.974	0	54.799	34828	163.3
3	-1	2	332.00	330.673	0	54.799	44137	185.0
4	-4	-10	332.00	329.828	0	54.799	11776	94.9
6	-4	-8	332.00	330.518	0	54.799	7593.9	76.0
4	-2	0	330.29	330.292	0	54.799	24404	137.3
8	2	10	332.00	330.583	0	54.799	4996.3	61.3
5	-4	-9	332.00	330.521	0	54.799	8041.4	78.3
5	-2	1	332.00	329.829	0	54.799	14447	105.3
0	-3	-15	332.00	330.080	0	54.799	7595.4	75.4
2	-3	-7	332.00	328.760	0	54.799	11586	94.0
3	-4	-11	332.00	328.150	0	54.799	5717.4	65.6
8	-1	5	332.00	329.325	0	54.799	4011.2	54.9
4	-4	-8	332.00	327.048	0	54.799	5317.0	63.3
4	-1	3	332.00	330.688	0	54.799	9151.5	83.6
4	-3	-3	332.00	327.669	0	54.799	5031.8	61.8
9	-2	3	332.00	327.358	0	54.799	958.73	26.6
3	-2	-1	332.00	329.744	0	54.799	1318.9	31.3
8	-4	-8	332.00	331.250	0	54.799	4544.1	58.8
9	-1	4	332.00	331.550	0	54.799	723.33	23.1
5	-3	-2	332.00	327.950	0	54.799	1937.1	38.1
6	1	5	332.00	327.950	0	54.799	183.62	11.6

b. Data Collection

The experiment consists of the collection of reflections as reciprocal lattice points pass through the sphere of reflection[31]. The goal of data collection is to collect as many reflections as possible in order to find the unique electron distribution in the unit cell that produces a calculated diffraction pattern that matches the observed intensities as closely as possible. The analysis is often simplified by the presence of symmetry in the unit cell. This reduces the problem to finding the density in the asymmetric unit only. The total electron density of the unit cell is then generated by the space-group symmetry operations, which can be deduced from the diffraction symmetry. In summary, X-rays will be scattered by crystals only in discreet directions. The locations of these directions are determined by the orientation of the crystal, the unit cell dimensions, and the wavelength of the x-rays.

The required components for the experiment are a crystal, a detector, an X-ray source with shutter, and a goniometer to orient and rotate the crystal. The SMART program controls the detector, the goniometer, and the shutter to create a series of images (frames) at specified goniometer setting angles. Each exposure is created by a simultaneous opening of the shutter and the rotation of the crystal by a small amount. At the end of the exposure, the shutter closes, the integrated counts collected in the

detector are transferred to computer memory or hard disk for subsequent processing, and the detector memory is cleared. The cycle is then repeated for the next image.

It is important to synchronize the shutter with the crystal rotation so that adjacent images truly reflect adjacent portions of reciprocal space. The detector should be positioned so that its entire active area is used, yet so diffraction spots do not overlap and the highest possible resolution is obtained.

Detectors generally do not have a uniform response over the entire active area. They also will distort the diffraction pattern in some geometric way. Spatial distortions and inhomogeneity of response must be corrected and defective pixels flagged, so that pixels can be mapped accurately to the actual location. Nonuniformity of response might arise with a detector because of variation in phosphor thickness, fiber-optic taper properties, pixel area, paths through windows, and so on.

At the University of New Orleans, the Chemistry Department has the state-of-the-art instrumentation for single crystal structure determination. The Bruker AXS SMART CCD diffractometer system is equipped with a SMART 1000 CCD detector, which has features for obtaining quality data sets in a fraction of the time required by instruments without CCD detectors. The Bruker CCD detector has a pixel size of 120 microns and contains 512 X 512 pixels. Physically, the SMART detector subsystem consists of four components in addition to the PC used to run the SMART data collection program. Provided is a picture of the detector system. First is the detector itself, mounted on the goniometer dovetail. Next is the Camera Electronics Unit (CEU) (not shown). The CEU

digitizes the analog signal from the detector and controls detector gain, exposure times, and other parameters under direction of the computer. Third is the PC interface card, which is located in the PC and connects the PC to the CEU.

Fourth is the refrigeration unit that cools the CCD to its operating temperature, which is -55ºC. Cooling the CCD chip reduces electronic noise. If the CCD temperature is above 0ºC the liquid crystal display (LCD), located on the front of the CEU, shows "WARM." Acquiring images while the detector is warm will not hurt the detector, but one will not be able to obtain good calibrations or x-ray data until it has cooled. The CEU has a built in thermal cutoff at 45ºC for extreme situations. The phosphor screen has a "built-in" miniature intensifier coupled to a fiber-optic taper that is connected to the CCD chip. Whereas traditional diffractometers use point detectors, which measure the intensity of each diffracted beam individually in sequence, the CCD detector records the diffraction pattern over a large area of reciprocal space simultaneously. The image is then stored digitally as a "frame" of diffracted intensity information. Many frames are collected as the crystal is rotated in a series of small steps. The frames are analyzed, and the intensity of each individual "reflection" can be determined. No predetermined information about the sample is needed to collect the data and solve the crystal structure, which is a major advantage.

c. Integration of Intensity

The result from data collection is a set of consistently measured, indexed intensities for as many of the reflections as possible. The criterion for finding peaks is

that a peak is represented by a set of contiguous pixels that are significantly above the local background. The raw intensities are processed to remove the contribution from background scattering and correct for certain geometric factors affecting the intensities. For data integration we use SAINT, a program by Bruker AXS, for integrating frames, applying Lorentz and polarization corrections, scaling, filtering, sorting and merging of reflections[32]. SAINT reads the raw data files (frames) along with the *.p4p file containing other crystal information. In order to integrate the data, the positions of all of the Bragg reflections must be accurately determined. The *.p4p file is used to determine initial positions, and updated cell dimensions and crystal orientations are determined by least-squares refinement during processing of the full data set.

For the integration each peak, a small volume of pixels is gathered from the current image as well as those before and after it to create a 3-D “shoebox” of pixels. The shoebox analogy is used since the length, width, and height can all be different, just as with shoeboxes. If the crystal orientation has been accurately determined, the peak will be located at the center of the shoebox, with the background on all six sides. A total intensity is calculated by summing the number of counts at each pixel inside the box, and a background calculated by summing the number of counts at each pixel on the side of the box. A net intensity is calculated by subtracting the background intensity from the total intensity.

The standard deviation in a raw intensity count is given by:

$$\sigma_I = N^{1/2} \qquad (2.8)$$

where N represents the number of counts. The standard deviation in the net intensity is given by:

$$\sigma_{net} = [(\sigma_{peak})^2 + (\sigma_{background})^2]^{1/2} \qquad (2.9)$$

The output of the integration is a *.raw file containing HKL indices, the intensity and its standard deviation, followed by the observed and calculated profile of the X,Y, and Z projections for each reflection.

With the output files from SAINT, the next step is to prepare for and create the files necessary for determination of the structure using the computer program XPREP. XPREP is used to determine the space group (symmetry), specify the unit cell contents, perform absorption corrections, and scale and merge data sets, etc. XPREP reads the raw data files from SAINT and the *.p4p file from SMART, and writes the crystal data file *.ins and reflection data file *.hkl to be used by later programs. The software shows the current crystal information and allows the user to choose from several options – the choice of unit cell, any of the Bravais lattice types, and any of the 230 possible space groups. Space groups specify the symmetry operations that are present in the crystal structure. In XPREP, usually the default option – the computer's best guess – is most likely correct for determining the space group and crystal lattice type. In this study, the monoclinic space group C2/c was selected based on the intensities of certain classes of systematically absent reflections, and the choice was subsequently confirmed by the successful determination of the structure.

Table 2.3 Summary of Preliminary Crystallographic Data

Original cell in Angstroms and degrees:

10.188 7.699 24.969 90.00 91.38 90.00

26375 Reflections read from file mlt18m.hkl; mean (I/sigma) = 5.51

Lattice exceptions: P A B C I F Obv Rev All

N (total) = 0 13195 13193 13194 13162 19791 17598 17590 26375

N (int>3sigma) = 0 3045 3158 3067 86 4635 4154 4166 6198

Mean intensity = 0.0 27.8 29.9 24.3 1.4 27.3 29.1 29.8 28.6

Mean int/sigma = 0.0 5.6 6.0 5.2 0.4 5.6 5.7 5.6 5.7

Lattice type: I chosen Volume: 1957.86

SEARCH FOR HIGHER METRIC SYMMETRY

Option A: FOM = 0.000 deg. MONOCLINIC C-lattice R(int) = 0.047 [9261]

Cell: 26.738 7.699 10.188 90.00 111.01 90.00 Volume: 1957.86

Matrix: 1.0000 0.0000 1.0000 0.0000 1.0000 0.0000 -1.0000 0.0000 0.0000

Option B: FOM = 0.000 deg. MONOCLINIC I-lattice R(int) = 0.047 [9261]

Cell: 10.188 7.699 24.969 90.00 91.38 90.00 Volume: 1957.86

Matrix: 1.0000 0.0000 0.0000 0.0000 1.0000 0.0000 0.0000 0.0000 1.0000

Option A selected

SPACE GROUP DETERMINATION

Lattice exceptions: P A B C I F Obv Rev All

N (total) = 0 13194 13166 13162 13195 19761 17600 17573 26375

N (int>3sigma) = 0 3067 3051 86 3045 3102 4158 4128 6198

Mean intensity = 0.0 24.3 24.3 1.4 27.8 16.7 29.3 29.2 28.6

Mean int/sigma = 0.0 5.2 5.2 0.4 5.6 3.6 5.7 5.7 5.7

Crystal system M and Lattice type C selected

Mean |E*E-1| = 0.938 [expected .968 centrosym and .736 non-centrosym]

Option Space Group No. Type Axes CSD R(int) N(eq) Syst. Abs. CFOM

[A] C2/c # 15 centro 1 3696 0.047 9261 0.7 / 5.7 2.14

[B] Cc # 9 non-cen 1 566 0.047 9261 0.7 / 5.7 6.27

Option [A] chosen

Determination of unit-cell contents

Formula: C20,H32,N2,O4

Formula weight = 349.44

Tentative Z (number of formula units/cell) = 4.0 giving rho = 1.186,

non-H atomic volume = 19.6 and following cell contents and analysis:

C 76.00 65.30 % H 116.00 8.37 %

N 8.00 8.02 % O 16.00 18.31 %

F(000) = 756.0 Mo-K(alpha) radiation Mu (mm-1) = 0.08

File mlt18.ins set up as follows:

TITL mlt18 in C2/c

CELL 0.71073 26.7383 7.6990 10.1877 90.000 111.005 90.000

ZERR 4.00 0.0087 0.0025 0.0033 0.000 0.008 0.000

LATT 7

SYMM -X, Y, 0.5-Z

SFAC C H N O

UNIT 76 116 8 16

TEMP -123

TREF

HKLF 4

END

26375 Reflections written to new reflection file mlt18.hkl

d. Creation of a Trial Structure

SHELXTL is an integrated system of computer programs for the determination and refinement of crystal structures using diffraction data, and provides simple steps for publication of the results. The program **XS** is used to generate trial structure solutions by calculating the phases of a subset of the hkl reflections from the SAINT output file. The program uses a number of different methods to try and guess the phases, and from them the identity and location of most atoms in the crystal. Hydrogen atoms are not usually found using this program. If **XS** is successful, then the trial structure generated may be examined by **XP,** a program for the visualization and editing of molecular structures**.** After a trial structure has been created, subsequent refinement cycles by **XL,** a least-square refinement program, and **XP** will eventually lead to finding all of the atoms. The two most common approaches used by **XS** to determine phases are direct methods and Patterson methods. Since the structure being investigated contains only small atomic number atoms, direct methods is the proper choice. As mentioned previously, the Direct methods approach is based on statistical analyses of the intensities of the reflections to find the most probable phase relationships. Remember, the phases cannot be determined experimentally and have to be calculated and combined with the experimentally determined amplitudes to give an electron density map. The Direct methods solutions from **XS** yield a list of positions called Qs. These Qs are peaks of normalized electron density found in the calculated E- map.

Structure factors are calculated using the equation:

$$F_{hkl} = \Sigma f_j \exp [-B_j (\sin\theta / \lambda)^2] \exp [2\pi i (hx_j + ky_j + lz_j)] \quad (2.10)$$

where f_j is the atomic scattering factor, B_j is the isotropic temperature factor. To normalize the structure factors, it is assumed that all of the atoms have the same isotropic temperature factor B and that they behave as point scatterers. The magnitude of a unitary structure factor, U_{hkl}, is found by dividing the structure factor of a point scatterer by the total number of electrons in the unit cell:

$$U_{hkl} = F_{hkl}, / \exp [-B (\sin\theta / \lambda)^2] \Sigma f_j \qquad (2.11)$$

Thus the normalized structure factor, E_{hkl}, is calculated as:

$$E_{hkl} = (U_{hkl}^2 / \langle U_{hkl}^2 \rangle)^{1/2} . \qquad (2.12)$$

Once all of the structure factors have been normalized, one can attempt to solve the phase problem. Since the crystal is centrosymmetric, the phase of a reflection is either $\pm$

that we have a good trial structure solution. XS writes the structure solution in the form of crystal data plus an atom list to the file *.res and a listing file *.lst.

e. Refinement of Crystal Structure

During the final stages of structure determination we try to improve the electron-density map generated by **XS**, by interpreting the map to produce an atomic model of the unit-cell contents, and refining the model to optimize its agreement with the original reflection intensities. The **XP** program is a graphical interface between the data calculated by **XS** or **XL**. This program will display the electron density as three dimensional contours and facilitates the ability to build the molecular model. It basically converts the ***.res file - which is an ASCII file containing the space group, symmetry, etc., data, and the atomic positions and displacement parameters in the earlier model and any Q peaks (i.e., peaks in the calculated electron density map that are not yet accounted for) - calculated by the **XS** or **XL** run into an easier to use form. When using **XP**, there are various graphical molecule viewing subroutines (e.g., 'proj'), one for deleting and naming atoms using a graphical tool (i.e., 'pick'), and routines to view peak positions and intensities in a tabular format (i.e., 'info') and calculate bond lengths and angles (i.e., 'bang'). Every time you start it, **XP** takes the last *.res file and uses it to generate all this data. The results are saved to *.ins.

The starting model generated by **XS** and displayed interactively by **XP** is improved by least-squares refinement of the atomic coordinates. The SHELXTL program **XL** reads the files generated from **XP** and writes the new results to the file

*.res. This method entails building a list of atoms, each with a set of coordinates (x,y,z) to specify its location.

By adjusting the atomic coordinates we improve the agreement between x-ray scattering amplitudes calculated from the current model and the original measured amplitudes in the native data set. Besides atomic positions, other parameters are included in refinement. The temperature factor, B_j of each atom *j*, is a measure of how much the atom oscillates around the position in the model. We know that molecules in the solid state are not static, they do have some freedom of movement, and diffraction is affected by this variation in atomic position. In this case, we assign a temperature factor to each atom and include the factor among parameters in minimizing the sum of the squares of the differences between observed and calculated amplitudes. Another parameter included in refinement is the occupancy n_j for each atom *j*, a measure of the fraction of atom j that actually occupies the position specified in the model. Occupancy is important if there is more than one conformation of the structure observed or the crystal contains more than one atom type at a certain location. In this study, there is no evidence of disorder or partial occupancy, so all of the occupancy factors have been fixed at 1.0. During this stage of structure determination, the model is cycled between map interpretation and least-squares refinement, **XP** and **XL**, respectively.

Before the parameters of the atoms in the model can be refined, reflections are determined to be observed or unobserved. A reflection that has an intensity less than three times the estimated standard deviation in its intensity is classified as unobserved

and is given zero weight. All other reflections are given unit weight. During the process of a least-squares refinement, parameters are shifted in an attempt to minimize the difference between observed structure factors and those which are calculated based on the model. Structure factors $|\mathbf{F}_c|$ are calculated using the following equation:

$$|F_c| = G \cdot \Sigma\, n_j\, f_j \exp[2\pi i\,(hx_j + ky_j + lz_j)] \cdot \exp\,[-B_j\,(\sin\theta)/\lambda]^2 \qquad (2.13)$$

where G is the overall scale factor, n_j is the occupancy of atom j, x_j, x_j, and z_j are the atomic coordinates. The second exponential term shows that the effect of B_j on the structure factor depends on the angle of the reflection [(sin θ)/λ].

Once the phases of each reflection has been estimated and a starting model has been proposed, the parameters of the model need to be refined in order to make the calculated structure factors better fit the observed ones. Similar to a least-squares fit of data to a straight line, we want to select atom positions that minimize the squared of differences between corresponding $|\mathbf{F_c}|$s and $|\mathbf{F_o}|$s.

The iterations of least-squares refinement are repeated until the parameters no longer shift significantly. Similar to a least-squares fit of data to a straight line, we want to select atom positions that minimize the squared of differences between the observed and calculated amplitudes for each reflection hkl. The quantity that is minimized is S, the sum of squares of differences between observed and calculated structure factors:

$$S = \Sigma\, w_{(hkl)}\,(\,|F_o| - |F_c|\,)^2 \qquad (2.14)$$

Depending on the relative number of parameters to be refined and observed reflections, the system may be under- or overdetermined or it may yield an exact solution. Because of the relatively low accuracy of any individual measurement, the number of observations should exceed the number of parameters, ideally by a factor of 10 or more.

The crystal structure refinement is converged when the weighted sum of the differences between the observed and calculated structure factors does not decrease on subsequent cycles. The sum is taken over all reflections hkl currently in use. Each difference is weighted by the term $w_{(hkl)}$, a number that depends on the reliability of the corresponding measured intensity. A reflection which has been observed with a smaller standard deviation can be given greater weight in the refinement process. The weight assigned to each measurement is inversely proportional to the square of the estimated standard deviation.

It is important to note that least squares may or may not lead us to the correct structure, as the starting model parameters must be near the global minimum, the one conformation that will give the best agreement between calculated and observed. Otherwise, the refinement will converge into an incorrect local minimum from which it cannot escape. In order to avoid this problem and increase the probability of finding the global minimum, we added constraints and restraints on the model during refinement cycles. Constraints and restraints have proven to greatly increase the rate of convergence of crystallographic refinements. A constraint is a fixed value for a certain

parameter, such as our constraint of all atomic occupancies to a value of 1.0. A restraint takes the form of additional information that is not exact but is subject to a probability distribution, such as the condition that all bond lengths, bond angles, and anisotropic displacement parameters are within a specified range of values. Commands in XL such as **DELU** and **SIMU**, were used to perform these types of restrains.

Thus in minimizing Q, we are finding atom positions, temperature factors, and occupancies that simultaneously minimize differences between (1) observed and calculated reflection amplitudes, (2) model bond lengths and ideal bond lengths, and (3) model bond angles and ideal bond angles. As the refinement proceeds, some constraints and restraints are lifted, so that agreement with the original reflection intensities is given highest priority. The choice when to relax specific constraints and restraints is more experience and art than science. The output of a least-squares minimization cycle calculated by **XL** is shown below.

Table 2.4 Crystal Structure Refinement Data

```
TITL Mlt18m in C2/c
CELL 0.71073  10.1877   7.6990  24.9688  90.000  91.384  90.000
ZERR    4.00   0.0033   0.0025   0.0080   0.000   0.008   0.000
LATT  7
SYMM -X, Y, 0.5-Z
SFAC C H N O
UNIT 76 116 8 16
```

V = 1957.86 F(000) = 756.0 Mu = 0.08 mm-1 Cell Wt = 1397.77
Rho = 1.186

TEMP -123

L.S. 4
BOND
FMAP 2
PLAN 10

WGHT 0.100000
FVAR 0.09492

C1	1	0.680018	-0.001924	0.576514	11.00000	0.00001
N1	3	0.844309	-0.426902	0.655577	11.00000	0.04934
O1	4	0.670749	0.151471	0.611026	11.00000	0.01140
O2	4	0.563511	0.003397	0.550766	11.00000	0.00001
C3	1	0.621966	-0.498495	0.641735	11.00000	0.00001
C4	1	0.740528	-0.499666	0.611948	11.00000	0.00001
C8	1	0.760205	-0.316294	0.689903	11.00000	0.01660
C9	1	0.718657	-0.308062	0.578426	11.00000	0.02516
C11	1	0.668561	-0.148749	0.610133	11.00000	0.00001
C13	1	0.740709	-0.147431	0.664851	11.00000	0.01570
C16	1	0.558191	-0.203471	0.516653	11.00000	0.02449
C20	1	0.627403	-0.411632	0.694281	11.00000	0.02833

HKLF 4

Covalent radii and connectivity table for Mlt18m in C2/c

C 0.770
H 0.320
N 0.700
O 0.660

C1 - O2 C11 O1
N1 - C8 C4
O1 - C1
O2 - C1 C16
C3 - C4 C20
C4 - C3 N1 C9
C8 - C13 N1 C20
C9 - C11 C4
C11 - C1 C13 C9

C13 - C8 C11
C16 - O2
C20 - C3 C8

h	k	l	Fo^2	Sigma	Why rejected
0	0	7	30.97	1.16	observed but should be systematically absent
0	0	9	28.42	1.37	observed but should be systematically absent
0	0	13	33.77	1.89	observed but should be systematically absent
0	-1	0	3.12	0.32	observed but should be systematically absent
0	-1	-1	58.94	0.79	observed but should be systematically absent
0	-1	1	56.25	0.86	observed but should be systematically absent
0	-1	-2	1.26	0.31	observed but should be systematically absent
0	-1	2	1.93	0.40	observed but should be systematically absent
0	-1	3	2997.29	19.38	observed but should be systematically absent
0	1	3	3000.17	24.72	observed but should be systematically absent
0	-1	5	59.22	1.30	observed but should be systematically absent
0	1	5	68.59	1.71	observed but should be systematically absent
0	-1	7	96.12	1.70	observed but should be systematically absent
0	3	17	27.61	2.56	observed but should be systematically absent
0	-3	19	139.34	4.28	observed but should be systematically absent
0	-3	23	92.43	4.74	observed but should be systematically absent
0	-3	25	38.27	4.34	observed but should be systematically absent
0	-5	-1	156.79	2.94	observed but should be systematically absent
0	-5	1	154.44	3.05	observed but should be systematically absent
0	-5	-5	8.50	1.47	observed but should be systematically absent
0	-5	5	7.04	1.70	observed but should be systematically absent
0	-5	-7	109.00	2.58	observed but should be systematically absent

** etc. **

26375 Reflections read, of which 13567 rejected

-15 =< h =< 15, -11 =< k =< 11, -37 =< l =< 38, Max. 2-theta = 66.47

2712 Systematic absence violations

3713 Unique reflections, of which 0 suppressed

R(int) = 0.0575 R(sigma) = 0.1017 Friedel opposites merged

Maximum memory for data reduction = 887 / 43888
Unit-cell contents from UNIT instruction and atom list resp.

C	76.00	72.00
H	116.00	0.00
N	8.00	8.00
O	16.00	16.00

Least-squares cycle 1 Maximum vector length = 511 Memory required = 1008 / 74396

wR2 = 0.9176 before cycle 1 for 3713 data and 49 / 49 parameters

GooF = S = 5.875; Restrained GooF = 5.875 for 0 restraints

Weight = 1 / [sigma^2(Fo^2) + (0.1000 * P)^2 + 0.00 * P] where P = (Max (Fo^2, 0) + 2 * Fc^2) / 3

N	value	esd	shift/esd	parameter
1	0.08877	0.00342	-1.801	OSF
5	0.00001	0.00459	-5.278	U11 C1
17	0.00001	0.00377	-3.025	U11 O2
37	0.00001	0.00517	-3.554	U11 C11

Mean shift/esd = 0.547 Maximum = -5.278 for U11 C1

Max. shift = 0.024 A for C16 Max. dU =-0.005 for C13

Least-squares cycle 2 Maximum vector length = 511 Memory required = 1008 / 74396

wR2 = 0.9210 before cycle 2 for 3713 data and 49 / 49 parameters

GooF = S = 5.721; Restrained GooF = 5.721 for 0 restraints

Weight = 1 / [sigma^2(Fo^2) + (0.1000 * P)^2 + 0.00 * P] where P = (Max (Fo^2, 0) + 2 * Fc^2) / 3

N	value	esd	shift/esd	parameter
1	0.07962	0.00330	-2.775	OSF
5	0.00001	0.00492	-7.152	U11 C1
17	0.00001	0.00402	-4.261	U11 O2
21	0.00001	0.00504	-3.349	U11 C3
25	0.00001	0.00497	-3.285	U11 C4
37	0.00001	0.00548	-5.030	U11 C11

Mean shift/esd = 0.710 Maximum = -7.152 for U11 C1

Max. shift = 0.029 A for N1 Max. dU =-0.006 for C20

Least-squares cycle 3 Maximum vector length = 511 Memory required = 1008 / 74396

wR2 = 0.9270 before cycle 3 for 3713 data and 49 / 49 parameters

GooF = S = 5.469; Restrained GooF = 5.469 for 0 restraints

Weight = 1 / [sigma^2(Fo^2) + (0.1000 * P)^2 + 0.00 * P] where P = (Max (Fo^2, 0) + 2 * Fc^2) / 3

N	value	esd	shift/esd	parameter
1	0.07263	0.00313	-2.234	OSF
5	0.00001	0.00553	-7.478	U11 C1
17	0.00001	0.00449	-4.736	U11 O2
21	0.00001	0.00567	-3.615	U11 C3
25	0.00001	0.00561	-3.501	U11 C4
37	0.00001	0.00605	-5.602	U11 C11

Mean shift/esd = 0.722 Maximum = -7.478 for U11 C1

Max. shift = 0.033 A for N1 Max. dU =-0.006 for C20

Least-squares cycle 4 Maximum vector length = 511 Memory required = 1008 / 74396

wR2 = 0.9323 before cycle 4 for 3713 data and 49 / 49 parameters

GooF = S = 5.317; Restrained GooF = 5.317 for 0 restraints

Weight = 1 / [sigma^2(Fo^2) + (0.1000 * P)^2 + 0.00 * P] where P = (Max (Fo^2, 0) + 2 * Fc^2) / 3

N	value	esd	shift/esd	parameter
1	0.06670	0.00298	-1.989	OSF
5	0.00001	0.00616	-7.866	U11 C1
13	0.00001	0.00533	-1.122	U11 O1
17	0.00001	0.00498	-5.200	U11 O2
21	0.00001	0.00631	-3.955	U11 C3
25	0.00001	0.00625	-3.759	U11 C4
37	0.00001	0.00655	-6.155	U11 C11

Mean shift/esd = 0.752 Maximum = -7.866 for U11 C1

Max. shift = 0.038 A for N1 Max. dU =-0.006 for C20

No correlation matrix elements larger than 0.500

Mlt18m in C2/c

ATOM	x	y	z	sof	U11	U22	U33	U23	U13	U12	Ueq
C1	0.68013	-0.00140	0.57614	1.00000	0.00001	0.06136	0.00380	0.00389	0.00148	0.00000	0.00616
N1	0.84347	-0.44220	0.65655	1.00000	0.04327	0.08816	0.00533	0.00635	0.00198	0.00000	0.01380
O1	0.66976	0.15103	0.61097	1.00000	0.00001	0.04039	0.00245	0.00275	0.00093	0.00000	0.00533

O2 0.56189 0.00090 0.55066 1.00000 0.00001
0.04318 0.00276 0.00255 0.00103 0.00000 0.00498

C3 0.62275 -0.49721 0.64223 1.00000 0.00001
0.06082 0.00378 0.00385 0.00144 0.00000 0.00631

C4 0.73876 -0.49690 0.61168 1.00000 0.00001
0.06332 0.00382 0.00409 0.00149 0.00000 0.00625

C8 0.75974 -0.31330 0.68967 1.00000 0.00368
0.06393 0.00411 0.00432 0.00147 0.00000 0.00840

C9 0.72577 -0.30542 0.57920 1.00000 0.01406
0.07559 0.00464 0.00530 0.00171 0.00000 0.01079

C11 0.66956 -0.14858 0.60973 1.00000 0.00001
0.05947 0.00362 0.00406 0.00136 0.00000 0.00655

C13 0.73714 -0.14831 0.66534 1.00000 0.00001
0.05890 0.00359 0.00415 0.00132 0.00000 0.00692

C16 0.55352 -0.20843 0.51656 1.00000 0.00706
0.07399 0.00429 0.00471 0.00159 0.00000 0.00901

C20 0.62333 -0.41370 0.69402 1.00000 0.00568
0.07004 0.00436 0.00496 0.00157 0.00000 0.00899

Final Structure Factor Calculation for Mlt18m in C2/c

Total number of l.s. parameters = 49 Maximum vector length = 511 Memory required = 959 / 22995

wR2 = 0.9375 before cycle 5 for 3713 data and 0 / 49 parameters

GooF = S = 5.168; Restrained GooF = 5.168 for 0 restraints

Weight = 1 / [sigma^2(Fo^2) + (0.1000 * P)^2 + 0.00 * P] where P = (Max (Fo^2, 0) + 2 * Fc^2) / 3

R1 = 0.7155 for 1191 Fo > 4sig(Fo) and 0.7870 for all 3713 data
wR2 = 0.9375, GooF = S = 5.168, Restrained GooF = 5.168 for all data

Occupancy sum of asymmetric unit = 12.00 for non-hydrogen and 0.00 for hydrogen atoms

Analysis of variance for reflections employed in refinement K = Mean[Fo^2] / Mean[Fc^2] for group

Fc/Fc(max)	0.000	0.010	0.019	0.029	0.039	0.052	0.065	0.084	0.111	0.155	1.000
Number in group	400.	345.	369.	388.	371.	356.	381.	362.	367.	374.	
GooF	4.252	4.617	4.720	3.756	4.818	4.606	5.344	7.390	9.354	14.131	
K	322.276	67.091	19.042	7.959	7.858	5.498	3.892	4.348	5.624	8.026	

Resolution(A)	0.65	0.68	0.70	0.74	0.77	0.82	0.88	0.97	1.11	1.40	inf
Number in group	379.	366.	378.	368.	368.	367.	374.	369.	371.	373.	
GooF	1.761	1.753	2.860	3.255	3.442	4.198	5.920	9.006	10.858	14.178	
K	2.161	1.955	3.032	3.429	3.061	3.193	4.165	6.714	7.481	12.096	
R1	0.608	0.576	0.617	0.610	0.637	0.650	0.678	0.725	0.727	0.807	

Recommended weighting scheme: WGHT 0.2000 0.0000
Note that in most cases convergence will be faster if fixed weights (e.g. the default WGHT 0.1) are retained until the refinement is virtually complete, and only then should the above recommended values be used.

Most Disagreeable Reflections (* if suppressed or used for Rfree)

h	k	l	Fo^2	Fc^2	Delta(F^2)/esd	Fc/Fc(max)	Resolution(A)
1	1	12	35555.53	11.60	5.66	0.013	1.96
6	0	0	77052.59	130.14	5.63	0.044	1.70
2	4	2	23398.21	45.77	5.62	0.026	1.78
-2	0	12	91563.55	8.90	5.58	0.011	1.94
3	3	6	13618.88	17.29	5.54	0.016	1.83
5	1	2	93485.11	737.54	5.52	0.104	1.94
-4	0	2	89593.06	1516.50	5.50	0.150	2.51
3	5	2	59966.98	933.03	5.46	0.117	1.39
0	4	12	17006.13	58.45	5.44	0.029	1.41
-2	4	2	30802.43	216.53	5.44	0.057	1.78
-5	5	2	50317.16	648.03	5.44	0.098	1.22
2	0	6	49099.22	1005.13	5.41	0.122	3.18
0	4	6	29580.78	492.01	5.37	0.085	1.75
-1	5	10	17271.33	25.27	5.35	0.019	1.30
-3	5	10	49554.13	923.57	5.33	0.117	1.23
-3	1	10	31895.00	1.83	5.33	0.005	1.97
0	2	4	145078.06	4281.27	5.32	0.252	3.28
-1	1	8	94615.68	2547.90	5.28	0.194	2.80
1	5	4	44882.29	375.28	5.27	0.074	1.48
0	0	12	34526.05	3.45	5.26	0.007	2.08
0	2	0	293489.13	8207.92	5.26	0.348	3.85
-1	1	4	523129.22	10775.31	5.25	0.399	4.41
0	6	8	16676.40	53.70	5.22	0.028	1.19
4	0	6	356848.09	12406.28	5.19	0.428	2.15
1	1	4	317169.75	10485.25	5.18	0.394	4.35
5	5	2	31578.16	269.31	5.15	0.063	1.22
3	1	2	59770.84	175.78	5.15	0.051	3.00
-1	1	10	98727.14	2065.70	5.10	0.175	2.32
3	1	10	77893.05	1588.77	5.09	0.153	1.93
-3	5	6	78018.32	3492.31	5.05	0.227	1.33
2	2	2	41193.86	2002.37	5.00	0.172	2.97
-2	0	6	149959.06	1131.24	5.00	0.129	3.26
3	3	2	20685.02	302.48	5.00	0.067	2.02
7	3	2	22114.27	823.38	4.94	0.110	1.26
-3	3	2	6452.27	21.87	4.93	0.018	2.02
5	3	12	29893.44	1383.30	4.91	0.143	1.25

3	1	12	64958.45	2714.63	4.88	0.200	1.71
2	0	0	682889.38	33480.09	4.88	0.704	5.09
-4	0	10	42880.37	831.76	4.87	0.111	1.80
1	1	0	124198.96	6453.20	4.86	0.309	6.14
-5	1	12	41853.81	2375.58	4.86	0.187	1.45
-5	5	10	20059.84	429.64	4.84	0.080	1.11
0	2	8	19151.21	729.15	4.84	0.104	2.42
-2	0	4	1064014.88	67644.37	4.83	1.000	3.99
0	4	8	12974.24	538.24	4.81	0.089	1.64
-4	6	2	60792.90	3092.68	4.80	0.214	1.14
4	0	20	35475.11	123.72	4.78	0.043	1.11
1	3	6	343005.41	23973.87	4.74	0.595	2.13
3	1	4	52334.20	755.87	4.74	0.106	2.76
2	0	26	60870.78	2109.13	4.74	0.177	0.94

Bond lengths and angles

C1 -	Distance	Angles		
O2	1.3488 (0.0473)			
C11	1.4153 (0.0438)	102.13 (2.87)		
O1	1.4659 (0.0383)	101.13 (2.60)	106.37 (2.82)	
	C1 -	O2	C11	

N1 -	Distance	Angles	
C8	1.5590 (0.0591)		
C4	1.5850 (0.0618)	100.17 (3.50)	
	N1 -	C8	

O1 -	Distance	Angles	
C1	1.4659 (0.0383)		
	O1 -		

O2 -	Distance	Angles	
C1	1.3488 (0.0474)		
C16	1.8238 (0.0425)	103.88 (2.38)	
	O2 -	C1	

C3 -	Distance	Angles	
C4	1.4220 (0.0525)		
C20	1.4439 (0.0520)	119.61 (3.39)	
	C3 -	C4	

C4 - Distance Angles
C3 1.4220 (0.0524)
N1 1.5850 (0.0620) 100.09 (3.28)
C9 1.6863 (0.0523) 101.71 (2.84) 98.71 (3.02)
C4 - C3 N1

C8 - Distance Angles
C13 1.4245 (0.0451)
N1 1.5590 (0.0591) 115.23 (3.36)
C20 1.5962 (0.0546) 109.19 (3.35) 102.48 (3.21)
C8 - C13 N1

C9 - Distance Angles
C11 1.5454 (0.0532)
C4 1.6863 (0.0521) 118.13 (3.11)
C9 - C11

C11 - Distance Angles
C1 1.4153 (0.0439)
C13 1.5349 (0.0474) 119.74 (2.90)
C9 1.5454 (0.0531) 107.39 (2.91) 106.49 (2.95)
C11 - C1 C13

C13 - Distance Angles
C8 1.4245 (0.0450)
C11 1.5349 (0.0474) 116.68 (2.93)
C13 - C8

C16 - Distance Angles
O2 1.8238 (0.0424)
C16 -

C20 - Distance Angles
C3 1.4439 (0.0520)
C8 1.5962 (0.0546) 98.02 (3.17)
C20 - C3

FMAP and GRID set by program

FMAP 2 2 13
GRID -2.500 24 -2 2.500 1 2

R1 = 0.7691 for 3713 unique reflections after merging for Fourier

Electron density synthesis with coefficients Fo-Fc

Highest peak 23.25 at 0.4395 0.0006 0.4499 [0.02 A from O2]
Deepest hole -4.03 at 0.9200 0.0004 0.4123 [1.42 A from C3]

Mean = -0.01, Rms deviation from mean = 2.01 e/A^3,
Highest memory used = 2115 / 32539

The structure solution is done when the structural model is complete and the refinement has reached convergence. A refinement is considered convergence when the largest shift is any parameter is less than 10% of its estimated standard deviation. We can determine whether or not our determined structure is correct by comparing the measured structure-factor amplitudes $|F_{obs}|$ with amplitudes calculated for our model, and by producing a flat difference electron density map. Serious peaks or dips in the electron density map may indicate that all the atoms have not been found or labeled incorrectly. In calculating the new phases at each stage, we learn what intensities our current model, if correct, would yield. Thus, the measured F_{hkl}s and calculated F_{hkl}s should agree within the accuracy of the data. The most widely used measure of quality of a structure determination is the residual index, or R-factor:

$$R_{hkl} = \Sigma_{hkl} | F$$

A well-refined model exhibits rms deviations of no more than 0.02Å for bond lengths and 4° for bond angles.

Finally, what makes one a good crystallographer? In my opinion, nothing other than solving more and more crystal structures should produce a proficient crystallographer since each crystal is never the same, so the analysis is always brings different challenges. Experience in judging crystal quality, understanding the analysis of the data at certain stages of the intensity integration, and several cycles of refinements have proved to be invaluable, affording to find discrepancies and making adjustments early, which oftentimes delays determining the structure, or wasting time on a crystal which will probably not produce any results.

The crystal structure of ethylene glycol bis(tropane-3-carboxylate), and the published crystallographic information is presented below in tabular form. A clear, parallelepiped crystal with dimensions 0.26 x 0.08 x 0.03 mm was mounted on a Bruker AXS SMART CCD diffractometer system equipped with a SMART 1000 CCD detector, graphite monochromator for data collection at 150 K. Lattice parameters were determined from least-squares refinement of 12,836 reflections measured from $1.63° < 2\theta < 33.23°$ using MoKα radiation (wavelength of 0.71073 Å). The unit cell was found to have the following dimensions: a = 26.738(9) Å, b = 7.699(3) Å, c = 10.188(3) Å, α = 90°, β = 111.005(8)°, γ = 90°. The unit cell has a volume of 1957.9(11) $Å^3$, and belongs to space group C2/c, with integrated intensities measured in the range $-37 \leq h$

≤ 37, -11 ≤ k ≤ 11, -15 ≤ l ≤ 15. Four molecules were found per unit cell. The molecular formula is $C_{20}H_{32}N_2O_2$, with a molecular weight of 364.48u, and a density of 1.237 Mg/m^3.

Out of the 12836 refined, 3725 are symmetry related. The averages of the symmetry related reflections yielded R_{int} = 0.0575 based on the magnitude of observed structure factors. The structure was solved by direct methods, peaks corresponding to the non-hydrogen atoms were located in the E-map, hydrogen atoms were found during refinement. For the data set collected, the final agreement factors and goodness of were R1 = 0.0453, wR = 0.1083, and GOF = 0.897. Final atomic coordinates and anisotropic temperature factors are listed in Tables 2.6 and 2.8. Bond lengths and bond angles are listed in Table 2.7. The molecular structure is depicted in Figure 2.9.

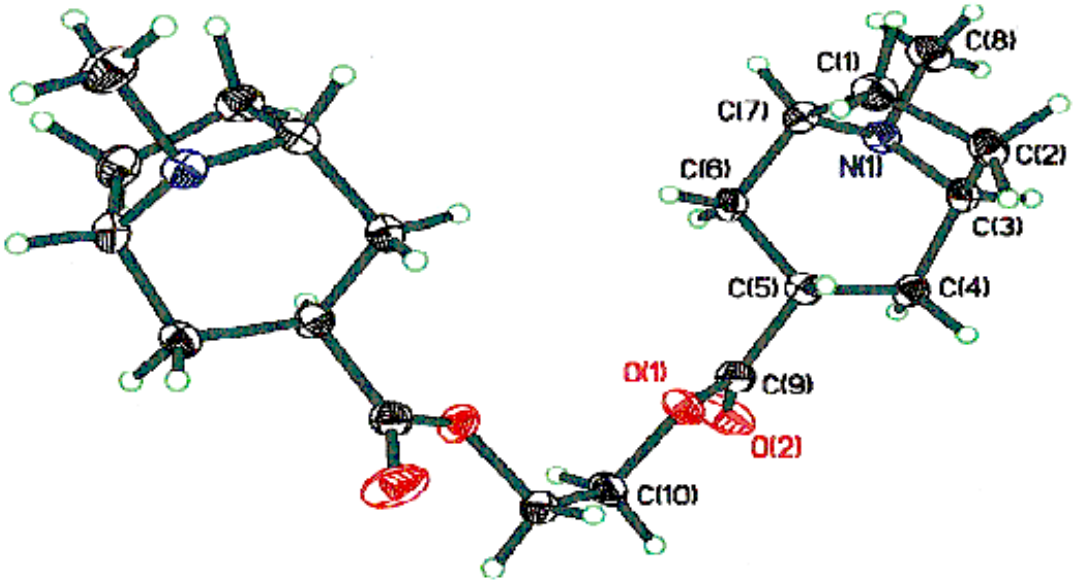


Figure 2.9. An ORTEP drawing of Ethylene Glycol BIS(tropane-3-carboxylate). Thermal ellipsoids are plotted at the 50% probability level.

Table 2.5. Published Results of Ethylene Glycol BIS(tropane-3-carboxylate)

Empirical formula	C_{10} H_{16} N O
Formula weight	364.48
Temperature	150(2) K
Wavelength	0.71073 Å
Crystal system, space group	Monoclinic, C2/c
Unit cell dimensions	a = 26.738(9) Å α = 90°. b = 7.699(3) Å β = 111.005(8)° c = 10.188(3) Å γ = 90°
Volume	1957.9(11) $Å^3$
Z, Calculated density	4, 1.237 Mg/m^3
Absorption coefficient	0.086 mm^{-1}
F(000)	792
Crystal size	0.26 x 0.08 x 0.03 mm
Theta range for data collection	1.63° to 33.23°
Limiting indices	$-37 \leq h \leq 37$, $-11 \leq k \leq 11$, $-15 \leq l \leq 15$
Reflections collected / unique	12836 / 3725 [R(int) = 0.0466]
Completeness to theta = 33.23°	98.7 %
Absorption correction	Empirical
Max. and min. transmission	1.000000 and 0.592338
Refinement method	Full-matrix least-squares on F^2

Data / restraints / parameters	3725 / 118 / 182
Goodness-of-fit on F^2	0.897
Final R indices [I>2sigma(I)]	R1 = 0.0453, wR2 = 0.1083
R indices (all data)	R1 = 0.0725, wR2 = 0.1238
Largest diff. peak and hole	0.313 and -0.320 e. $Å^{-3}$

Table 2.6. Fractional Atomic coordinates ($x10^4$)

Atom	*X*	*Y*	*Z*

Table 2.7. Bond lengths [Å] and angles

BONDED ATOMS	***BOND LENGTHS***

Table 2.8. Anisotropic displacement parameters ($Å^2$ x10^3) for mlt18.
The anisotropic displacement factor exponent takes the form:
$-2\ pi^2\ [\ h^2\ a^{*2}\ U11 + ... + 2\ h\ k\ a^*\ b^*\ U12\ ]$

ATOM	*U11*	*U22*	*U33*	*U23*	*U13*	*U12*

H8A	1571(6)	13979(19)	-255(17)	39(4)

CHAPTER III

CORRELATIONS BETWEEN STRUCTURAL PROPERTIES AND STRENGTH PROPERTIES OF COTTON FIBERS GROWN IN VARIOUS COUNTRIES

3.1 Introduction

Recently, deliberate mislabeling of the geographic origin of foreign merchandise and food products by manufacturers for profit has become increasingly common and is a growing concern for the United States government's enforcement of quotas. The "country of origin" is the country in which a product is wholly produced or manufactured (except for minor parts). The country of origin designation of import goods is used to determine whether they are subject to import prohibition, restriction, or they qualify for preferential duty admission[33]. Research on new methods to determine the country of origin of targeted commodities and food products is increasing rapidly[34,35,36]. These studies are particularly important in the cotton textile industry, as the unlawful mislabeling of cotton can introduce variability in the quality of products produced form the raw material, ultimately damaging the industry, and producing a negative economic impact on the global market and trading practices. According to the United States Cotton Standards Act[37], the sale, advertisement, and description of cotton must be regulated in order to protect the interest of the producers, merchandisers, processors, and consumers by proper and reliable classification. To the best of our knowledge, no research studies on methods for the determination of the country of origin of raw, unprocessed cotton have been reported.

Egypt produces some of the finest cottons in the world. Grown exclusively in the Nile River Valley, this land has excellent soil to grow extra long staple cotton. In general, the fibers of Egyptian cottons are longer, stronger, and finer than those of other cottons[38,39,40]. Products produced from Egyptian cotton are considered superior to those manufactured from other cottons, and thus command higher prices. Since Egyptian cotton is subject to a quota, it is of interest for government agencies to have a method for determining the country of origin, and to investigate the correlation of the structural properties of raw cottons with their country of origin. The ability to distinguish Egyptian raw cotton from cotton from other source countries would aid in the enforcement of this quota.

The differences of between Egyptian cottons and cottons of other origin are presumed to be genetic in nature, suggesting that Egyptian cottons might have a somewhat different crystalline structure from that of other cottons[41,42,43]. Environmental conditions, including the soil composition where the cotton is grown, may also influence the crystalline structure and chemical composition of the cotton. Described in this report are investigations to determine if Egyptian cotton and cotton from other source countries can be distinguished based on measurements of their X-ray diffraction patterns. It is proposed that by determining crystallinity, crystallite size, and orientation of the fibers of cottons from several countries, structural data will be obtained that correlates with physical strength, the standard of fiber quality. The hypothesis is that, if the structural characteristics of Egyptian cottons are sufficiently different from cottons grown in other

countries, then x-ray diffraction should produce fiber diffraction patterns which may be used to discriminate between those cottons.

X-ray fiber diffraction analysis of cotton fibers has been employed to determine percent crystallinity,[44,45,46] crystallite orientation,[47,48] and crystallite size.[49,50] Knowledge of relative variations in crystallinity, crystallite size, and orientation of crystallites to the fiber axis has been shown to be helpful in understanding fiber properties and chemical reactivity[51]. These structural characteristics of cotton fibers may reveal distinct differences between the cottons of various origins. In this study, cotton samples grown in Egypt and the U.S., as well as cottons from Australia, South Africa, Greece, China, India, and other countries have been analyzed using a synchrotron x-ray radiation source equipped with a sensitive CCD area detector. This is the first report in which cotton fibers have been analyzed in this manner.

3.2 Cotton Fiber Structure

Cotton is classified as a fiber and a food that is grown in over 80 countries, producing over 21.6 million tons worldwide for apparel and home furnishings[52]. A cotton fiber is actually the tubular outgrowth of a single cell on the epidermis of the cotton seed. The mature cotton fiber is a dead, hollow dried cell wall tubular structure, which is collapsed, shriveled, and twisted, giving the cotton fiber convolutions (twists)[53]. It is believed that cotton fibers grow in three distinct stages of development: elongation, secondary maturation and dehydration, inside a green capsule, the cotton boll. The fiber consists of three main parts: the primary wall, the secondary wall, and the lumen.

The secondary wall layers consist of fibrils of pure cellulose laid down spirally about the axis of the fiber.

The angle of the helix, which the cellulose fibrils make with respect to the fiber axis, is referred to as fibrillar orientation, one of the most important structural features of cotton fibers.

The composition of the cotton fiber is about 95% cellulose, and is almost wholly crystalline. In the crystalline regions of the cotton fiber, the molecular chains of cellulose lie parallel in three-dimensional arrangements of high geometrical order. In the amorphous regions of the fiber, the molecular chains are arranged in less ordered states. There are no sharp boundaries between the two regions, only in random areas along the fiber axis. The chains of cellulose molecules associate with each other by forming hydrogen bonds. They join together to form microfibrils also called crystallites. The microfibrils organize into macrofibrils, and the macrofibrils organize to form fibers[54]. Cotton cellulose is not a single crystal but rather a crystalline aggregate. The crystalline regions in cotton cellulose do not have sharp boundaries - they are interspersed with less crystalline areas and with some non-crystalline (amorphous) areas. In the long chains of cellulose molecules in cotton (sometimes thousands of glucose units long), some portions are in an orderly arrangement with respect to their neighbors while other portions lie in a disoriented, or random arrangement.

Since the x-ray diffraction pattern of every crystalline substance is characteristic and distinctive, x-ray techniques are a valuable tool in studies of molecular structure[55].

The crystal structure of a particular cotton species, variety, or sample determines many of its important physical properties. High strength and low extensibility (elongation) are associated with a high percentage of crystallinity and with relatively good orientation of the crystallites with respect to the fiber axis[56].

To measure cotton fiber strength, a bundle of cotton fibers are pulled with an increasing amount of force until the bundle breaks. The reading on the scale is expressed as breaking stress or force to break per linear density of the bundle, g/tex. The measurement of the mean orientation of the crystallites varies greatly among the varieties of cotton and it is also affected by the conditions under which the cotton was grown[57].

3.2 X-ray Analysis

a. Apparatus

For preliminary collection of X-ray fiber diffraction data, a 3-circle X-ray diffractometer (model Bruker SMART 1000, Bruker AXS Inc. Madison, WI), SMART 1000K CCD detector (detector temperature: –53.59°C, 120 micron pixel size in a 512X512 image), and a sealed tube Molybdenum Kα radiation source operated at a voltage of 45kV and current of 35mA with a graphite monochromator was used. For the final data collection of cotton fiber diffraction patterns, a normal-conducting electron storage ring producing synchrotron x-rays at 8.0425 keV with a beam size of 0.5mm was utilized. The diffraction patterns were recorded with a Mar CCD detector with a

2048X2048 resolution and a 78.838 micron pixel size, while the detector temperature was -79.60°C.

b. Sampling, Preparation, and Analysis

Several samples, representing different varieties of cottons from various countries, were obtained from the Bremen Cotton Exchange in Germany. The countries represented in the analysis were America (Giza 45, Giza 75, and US Pima), Asia, Australia, China, Egypt (Giza 70, Giza 75, and Giza 80), Greece, India, Peru, South Africa, Turkey, Uzbekistan, and Zimbabwe. Each bag of cotton (sample) was considered to be a representative of the bulk. Aliquots were collected from various locations in the bag to obtain a representative of the sample. Each aliquot weighed approximately 0.01-0.025mg. The aliquots taken from each sample were gently comb-separated by hand to obtain a small bundle of parallel fibers. These fibers were mounted on the x-ray diffractometer with the fibers parallel to the Φ axis, and perpendicular to the x-ray beam. The output of the detector was called a “frame”, and several frames were collected for each cotton sample.

Since a total of 16 different cottons were used in this study, it was important to determine an appropriate analysis time. In preliminary experiments, the frames were collected at 7hr, 5hr, 3hr, 1hr, and 30min intervals respectively on the same cotton

sample to determine the optimal exposure time. A 30min exposure time provided sufficient signal to noise ratios, and longer exposure times were unnecessary. On the other hand, using the synchrotron source, it was concluded that only a 3 minute scan was sufficient. Since the synchrotron data was found to be superior to data collected using the in-house Bruker AXS x-ray diffractometer, the synchrotron was employed to collect data on all of the samples analyzed.

Ten scans at room temperature were performed on each bundle of fibers oriented perpendicular to the x-ray beam. Each scan was collected at fixed omega, 2-theta (detector position) and chi angles, and a phi angle rotation of 30 degrees, recorded as an individual frame as seen in Figure 3.1. The data was corrected for air scatter (obtained by measuring a frame without a sample present), which removed most of the background from the diffraction pattern.

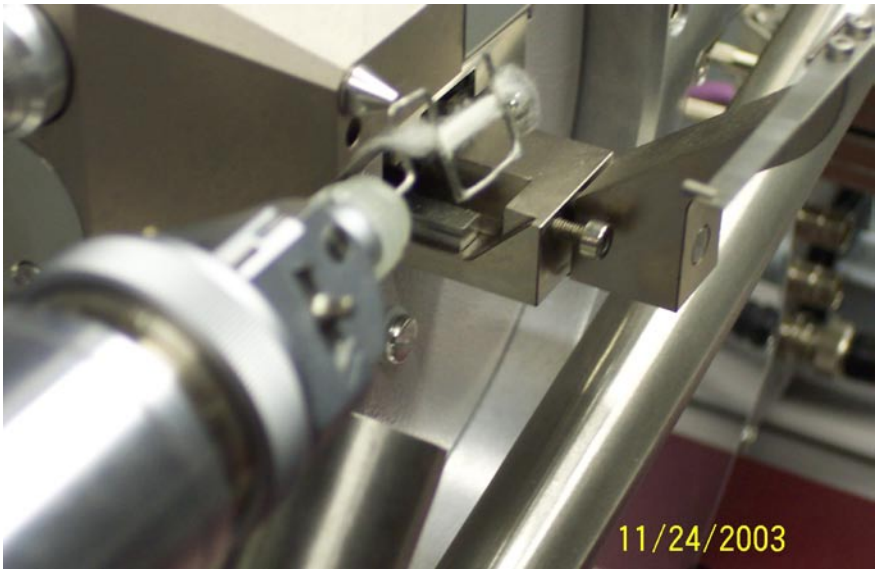


Figure 3.1. Cotton fibers aligned along the phi axis. The x-ray beam is shown to the left of the bundle. Photo taken at the CAMD LSU Facility in Baton Rouge, Louisiana.

In the analysis of fiber diffraction patterns obtained, two forms of integration were employed: χ and 2θ. The regions along the specified 2θ and χ directions were integrated and plotted as intensity versus the χ or 2θ angles. Diffraction intensities were calculated at 0.05° steps over the angular range 0° to 30° (2θ). The χ integration range was a full 360° for the chi integrations and approximately 345° for the 2-theta integrations. For integrating along either the 2θ or χ directions, Bresenham's algorithm[58] was used to determine the pixel count along that line. Each pixel's intensity was weighted and summed with the result normalized. All plots obtained from each sample were further analyzed using the Jade[59] software package to obtain the peak full width at half maximum (FWHM), d-spacings, 2-theta values, etc., through profile fitting using appropriate functions. A calculated diffraction pattern was used to compare with the resolved peaks. The cotton fiber characteristics that are of most interest are crystallinity, crystallite size and orientation. By obtaining these parameters, one can make correlations between the structure of the cotton fibers and its physical properties. The most intense peak, the 002 reflection, was used in the characterization of the cottons for the x-ray analysis. The amorphous regions and the crystalline region of the 002 peak were used in the percent crystallinity calculations. The FWHM of the 002 peak in the 2θ direction was used to derive the crystallite size estimate, and the FWHM of the 002 peak in the chi direction determines the crystallite orientation.

c. Calibration

The purpose of instrument calibration is to eliminate or reduce bias in an instrument's readings over the range for all continuous values[60]. For data collection, a x-ray reference standard with known values is measured with the instrument in question to validate the instrument's response and to confirm the instrument is working properly. For this x-ray analysis, NIST standard reference material 676 (Table 3.1) was used for calibration. The alumina (corundum) standard was placed in a 0.5mm capillary tube and exposed to x-rays at E = 8.980 keV. Immediately following the data collection, cotton samples were placed on the diffractometer using the same detector position. From the known d-spacing of alumina, accurate sample-to-detector distances were calculated, which were then used to calculate 2-theta values and d-spacings of the cotton samples. The broadening of all observed diffraction peaks can be characterized by the FWHM (Full Width at Half Maximum) value of the cotton fibers and the diffraction instrument. In order to determine the instrumental broadening from the synchrotron source, corundum was utilized since it does not exhibit significant sample broadening. Therefore, the broadening of the diffraction peaks was considered to be only due to the crystallite size of the cotton fibers (to be discussed later). After profile fitting of the standard, it appeared that the FWHM instrumental broadening is negligible. The instrumental FWHM curve was included in Jade© for crystallite size analysis of the cotton fibers.

Table 3.1. Results of calibration of the XRD instrument

Reflection	*Corundum Standard Nist 676*	*Corundum Standard Nist 676*	*Synchrotron CAMD*	*Synchrotron CAMD*	*Sample to Detector Distance (mm)*
	2θ°	*d(Á)*	*2θ°*	*d(Á)*	

3.3 Results

Table 3.2. Diffraction Measurements - 002 PEAK

SAMPLES	***2θ***	***Std. Dev.***	***d spacing***	***Std. Dev.***

Table 3.3. Diffraction Measurements - 002 peak

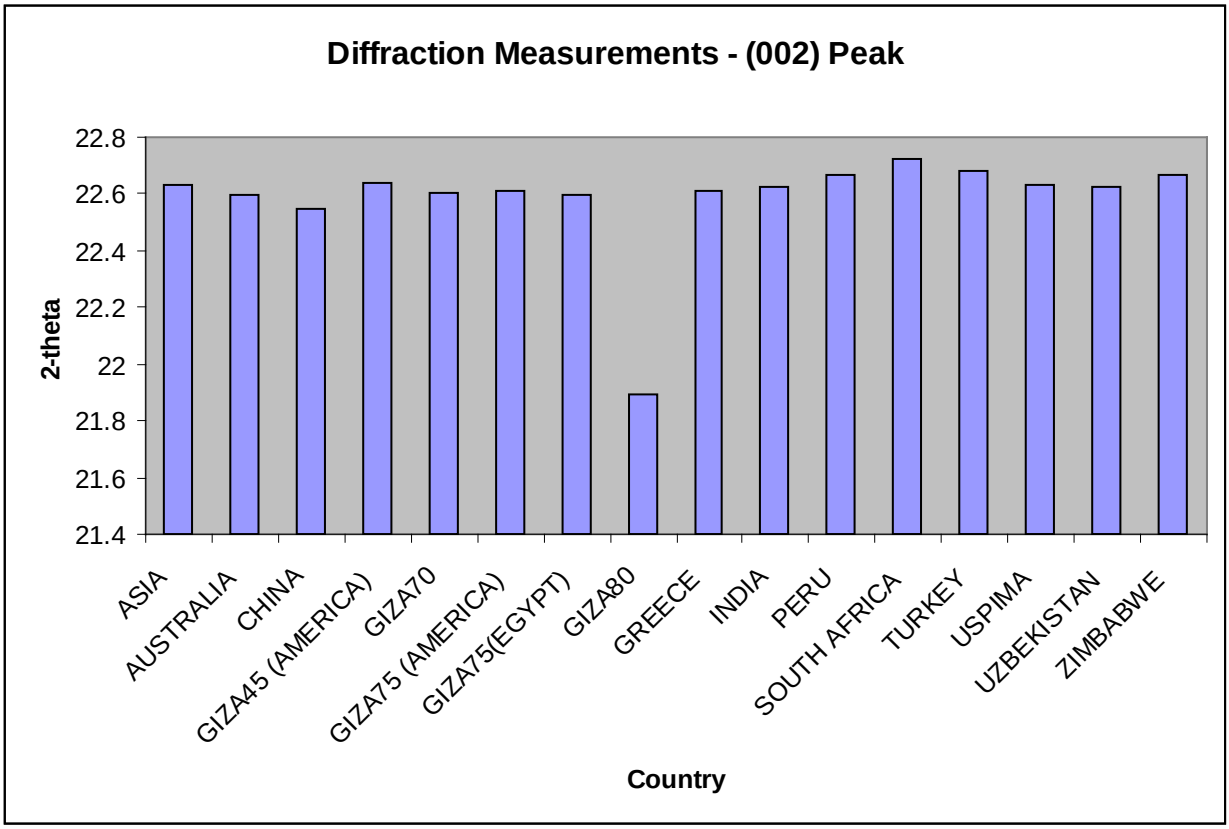


Table 3.4. Crystallinity Results

SAMPLES	***MEAN***	***STANDARD DEVIATION***

Table 3.5. Crystallinity Results

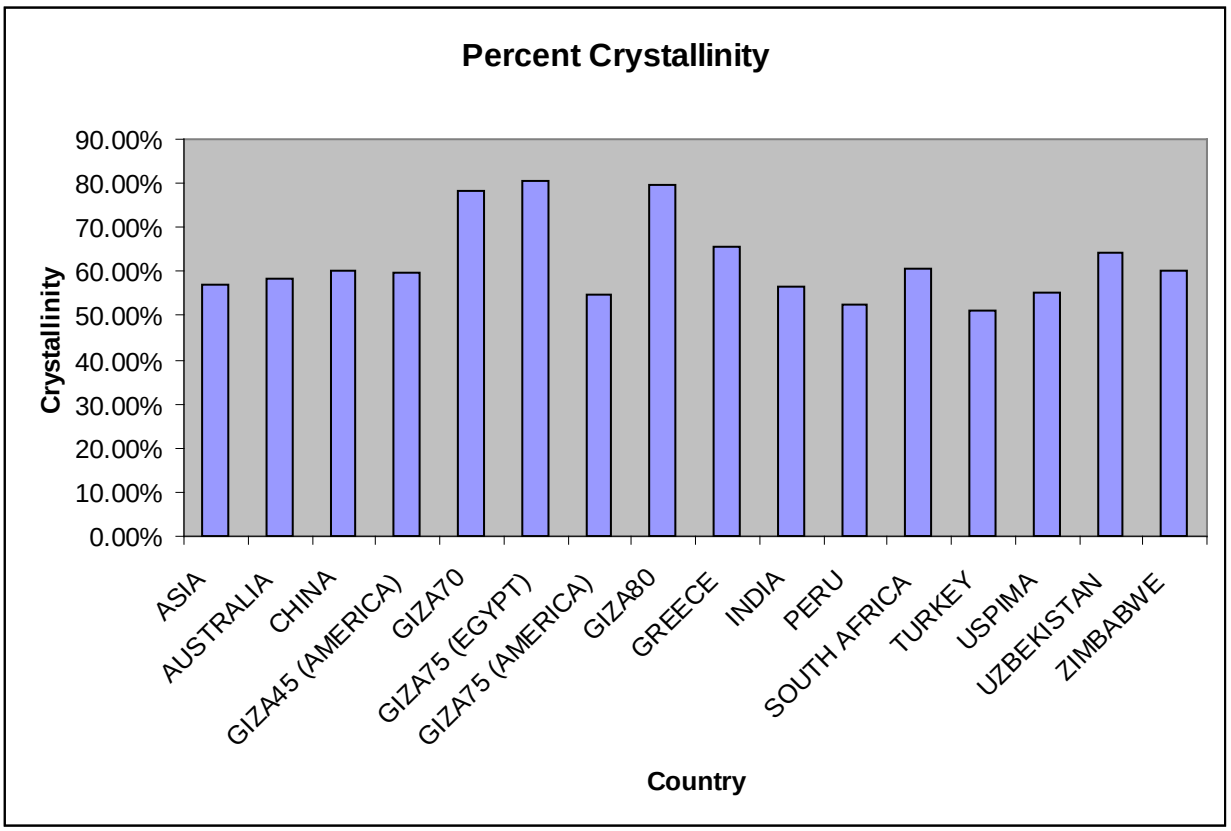


Table 3.6. Crystallite Size Measurements

SAMPLES	***FWHM (MEAN)***	***FWHM (STD. DEV)***	***CRYSTALLITE SIZE (Å)***

Table 3.7. Crystallite Size Estimates

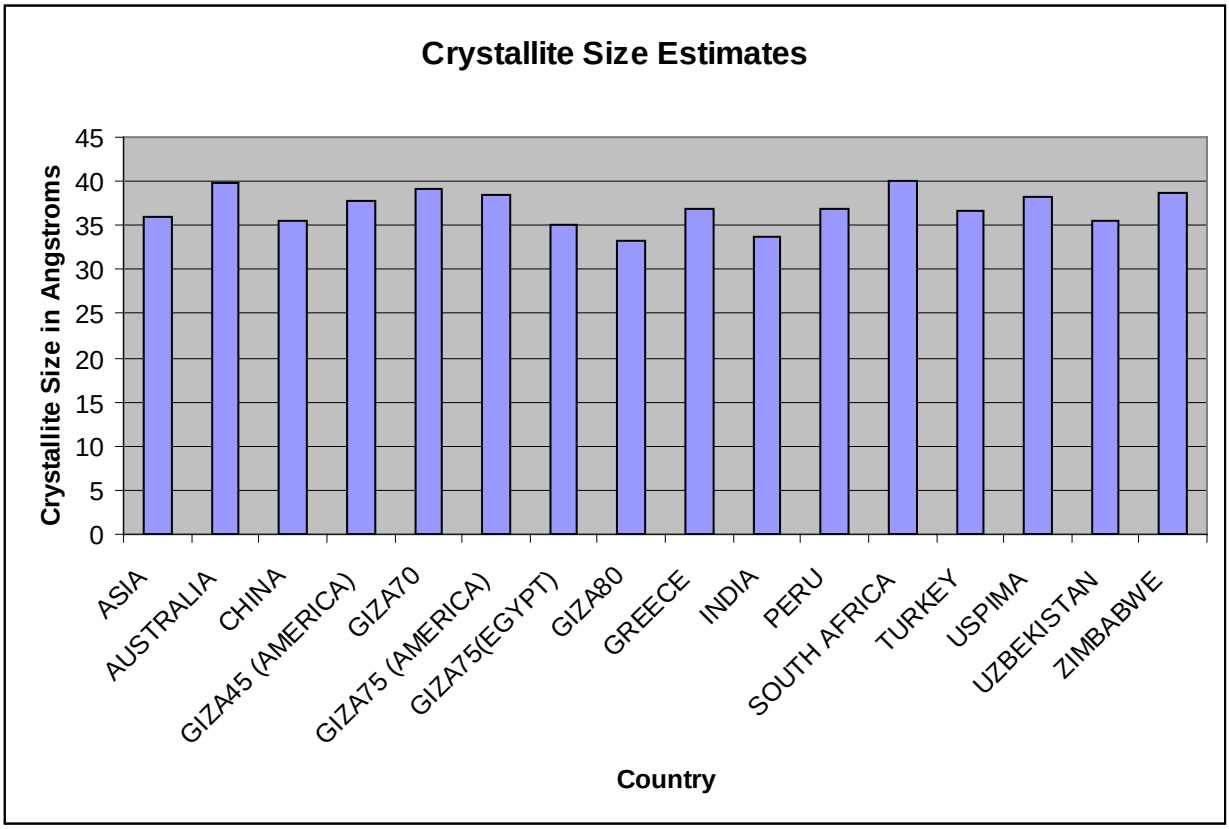


Table 3.8. Crystallite Orientation

SAMPLES	***FWHM (MEAN)***	***STANDARD DEVIATION***

Table 3.9. Crystallite Orientation

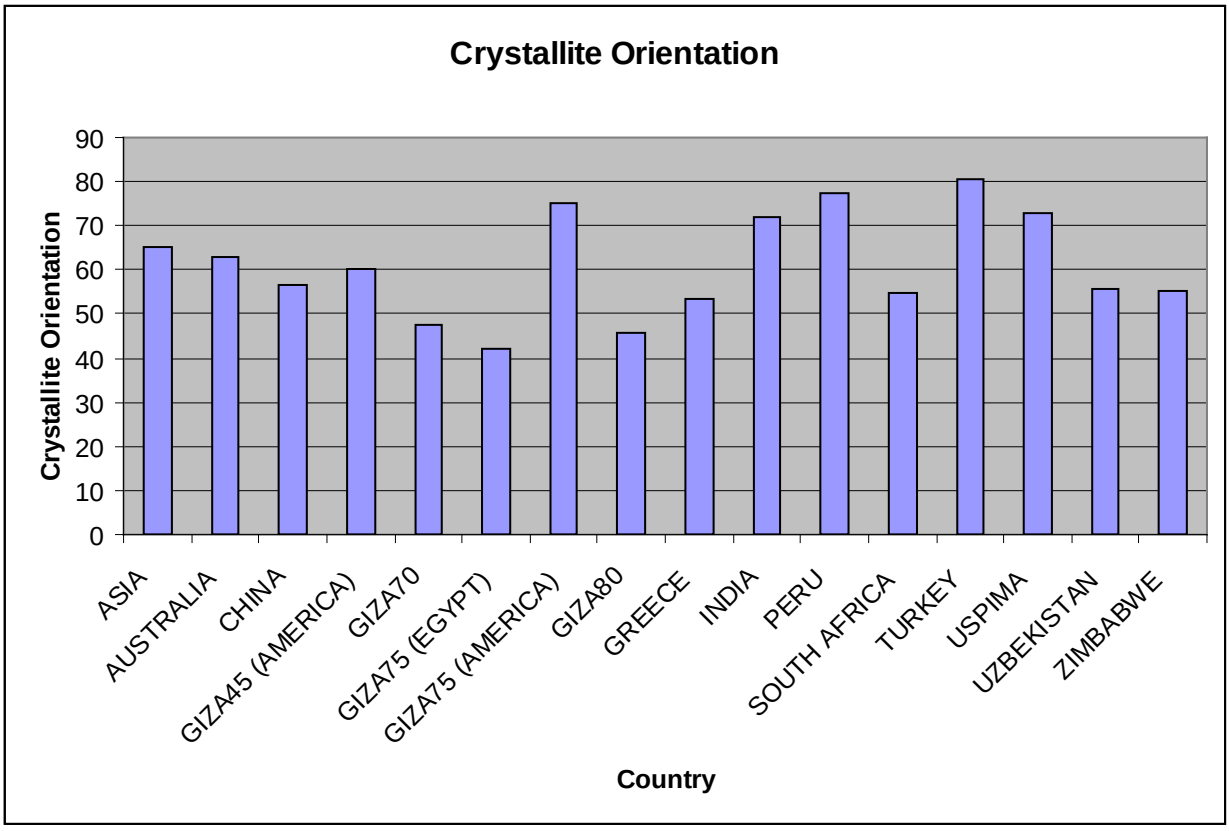
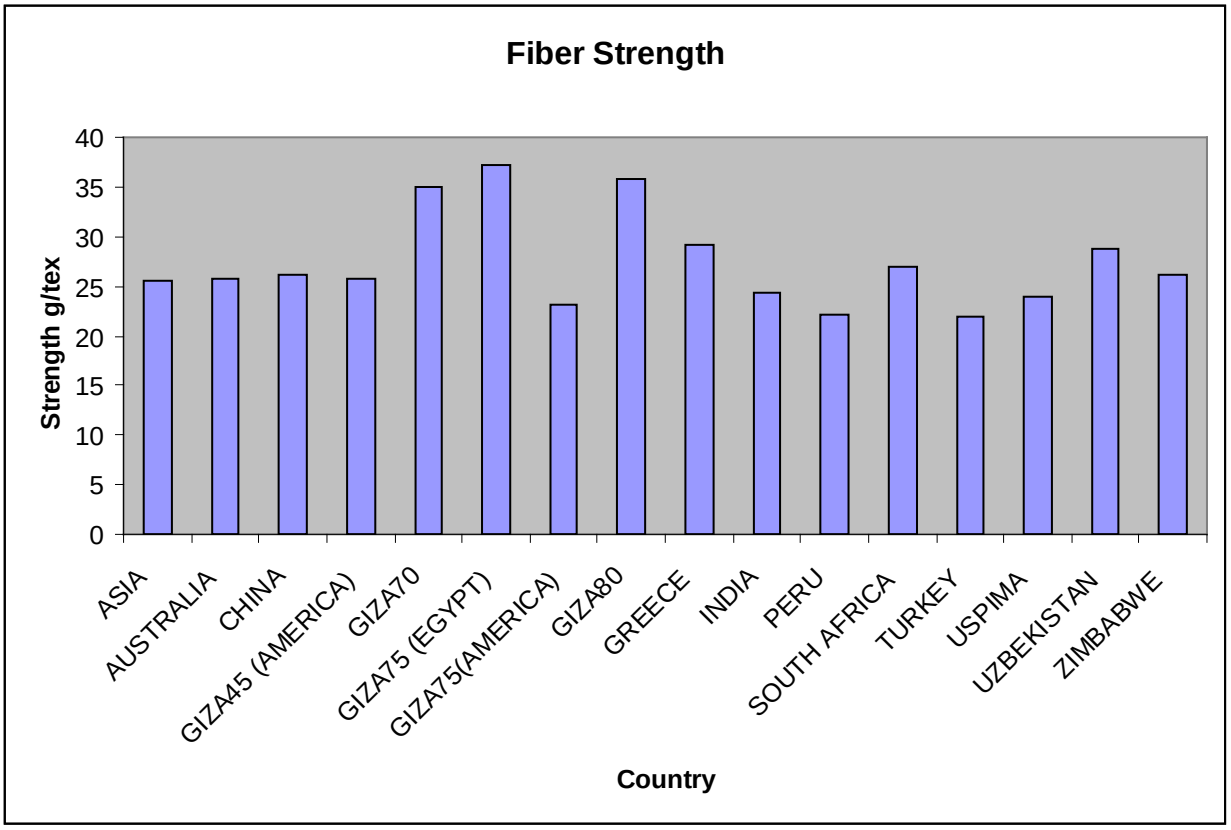


Table 3.10. Measured Strengths of Cotton Fibers

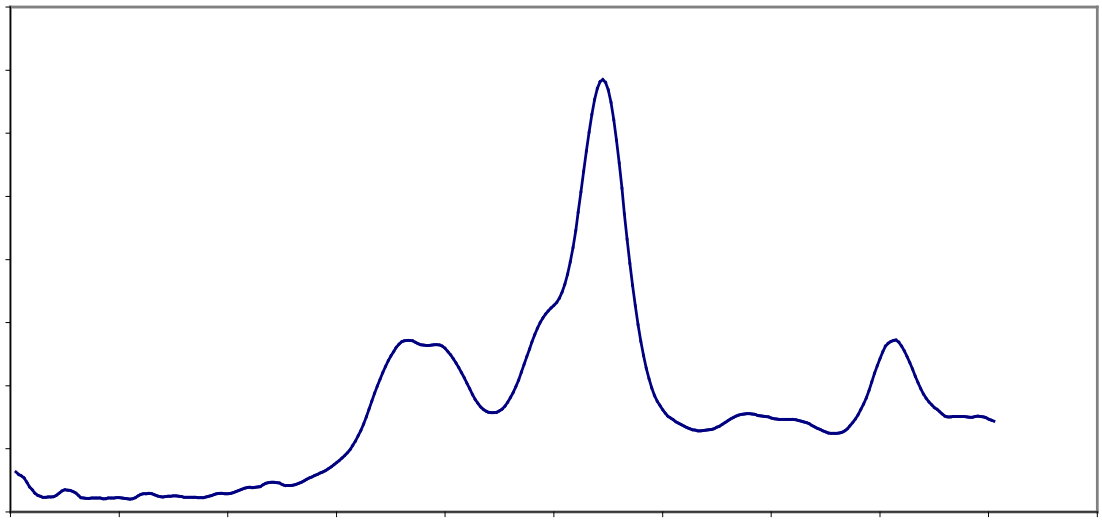
SAMPLE	***STRENGTH g/tex***

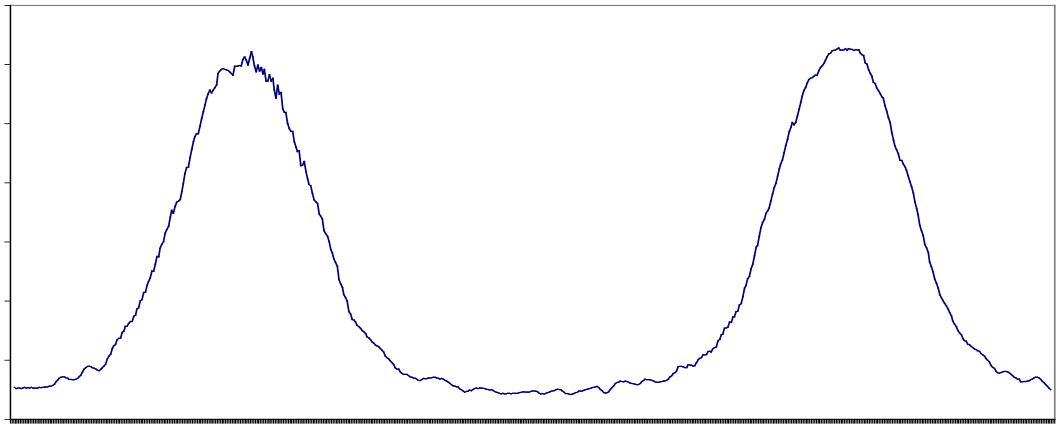
Table 3.11. Measured Fiber Strength



3.4 Discussion

The following figures illustrate typical cotton fiber diffraction patterns resulting from chi and 2-theta integrations.





The synchrotron diffraction patterns and integrations are typical of cotton fibers diffracting x-rays from the principal reflection planes of the cellulose crystallites: (Figure 3.3: left to right) 101, 10$\underline{1}$

3.5 Data Treatment

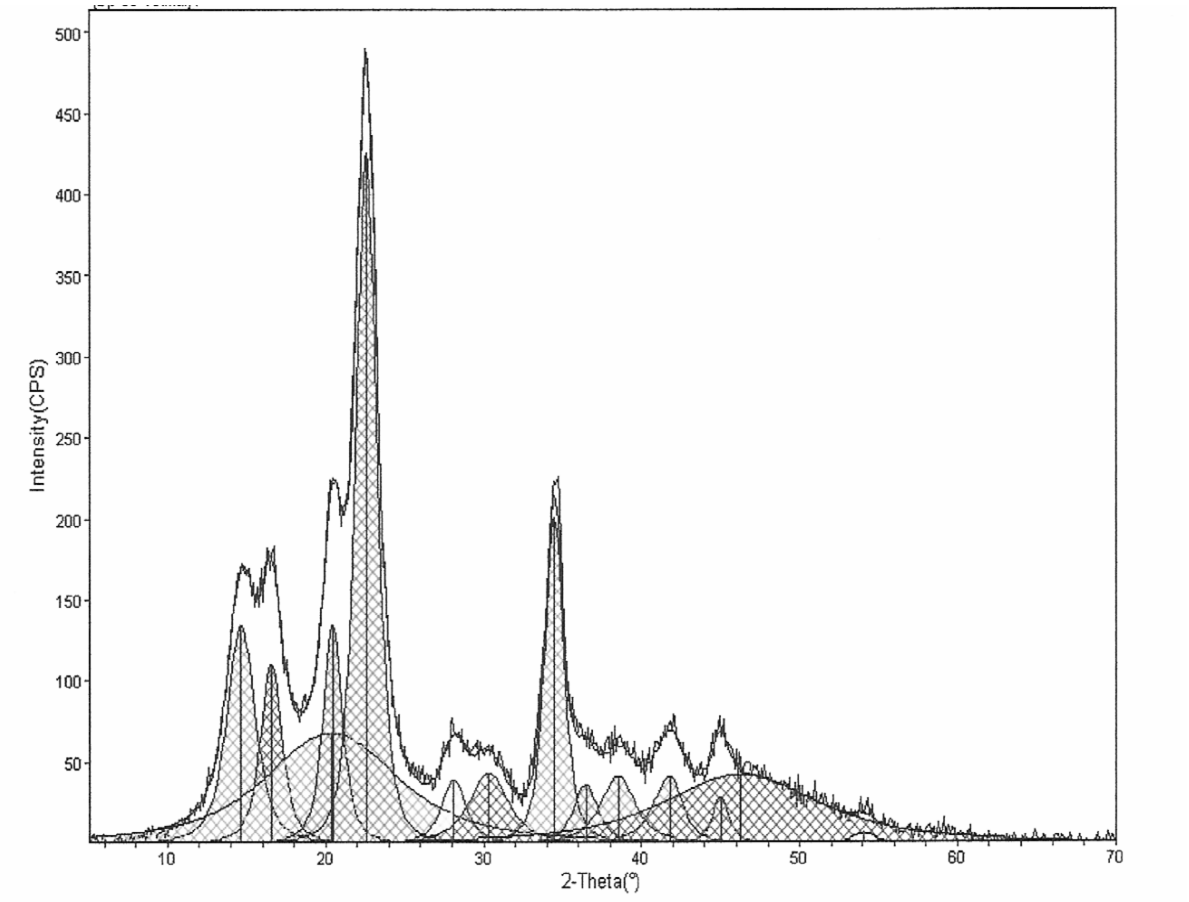
The direct analysis of diffraction profiles from cotton fibers was challenging because of strong peak overlap. To deconvolute the overlapping peaks and obtain peak shapes and positions, it was necessary to approximate diffraction profiles using suitable analytical functions and to perform least-squares iterations of refinable parameters. All diffraction patterns were corrected for absorption, air scatter, Compton scattering, polarization, and the Lorentz factor, and resolved into peaks. The measured diffraction profiles were analyzed in this manner using the computer program Jade. This program resolves multiple peak data into individual peaks and a background. It incorporates an iterative refinement procedure based on minimizing the following quantity:

$$R = \sum_{i=1}^{n} (Y_{(OBS)i} - Y_{(CAL)i})^2 \quad . \quad (3.1)$$

Each peak is represented by four parameters: the profile function, peak height, peak width, and peak position. The χ and 2θ integration results were analyzed by profile fitting to yield the positions and widths of individual diffraction peaks. Jade provides several profile functions to choose from: Gaussian, pseudo-Voigt, Pearson VII. The function which gave the best R-factor and goodness of fit for the corrected experimental data, the Pearson VII profile function, was selected. All of the major peaks described above were modeled with the Pearson VII shape function along with parameters

describing skewness, tail shape, and a linear background, which are refined together (Figure 3.4).

During the least-squares fitting, the refined parameters after each iteration cycle were output together with the R-factor (GOF), which should be below about 5% indicating a satisfactory model of the crystalline structure. Excellent agreement between experimental and calculated profiles were obtained, with R-factors ranging from 1% to 3%.


Intensity(CPS)
500
450
400
350
300
250
200
150
100
50
10
20
30
40
50
60
70
2-Theta(°)

3.6 Crystallinity Estimates

Crystallinity measurements in native cellulose materials have received attention for commercial applications because of the importance of cellulose as industrial raw material[63]. The percent crystallinity of each sample was obtained by using the Segal *et al*[64] empirical method in which the crystallinity was defined as the ratio of the intensity of x-ray scattering from the crystalline region of the cotton fiber to the intensity of the crystalline and amorphous regions combined (equation 3.2). The x-ray intensities are considered to be the integrated x-ray scattering under the resolved peaks, above the background.

$$\% \text{ Crystallinity} = 100 \times I_{cry} / [\, I_{cry} + I_{am} \,] \quad (3.2)$$

I_{am} is the total intensity of the amorphous regions of the fiber and I_{cry} is the total intensity diffracted from the crystalline component of the cellulose molecules.

As shown in the Table 3.4 above, the percent crystallinity estimates ranged from 52-80%, revealing significant differences between the cottons, particularly the Egyptian cottons. For example, the Egyptian cotton Giza 75 is 36% more crystalline than cotton originally from Turkey. Giza75 grown in Egypt is 32% more crystalline than Giza75 grown in America. The estimated standard deviations were derived by comparison of repeated scans were in the range of 0.437-1.56%. The random nature of x-ray

scattering may be one source in the variation between repeated scans. Another factor may be the amount of sample in the x-ray beam.

As the sample was rotated in 10 degree increments totaling 60 degrees phi, the amount of sample in the beam could change, leading to variable crystallinity estimates for the same sample. One way to avoid this would be to improve sample preparation and mounting in order to get a uniform sample throughout the x-ray beam.

The correlations between the fiber strength and the percent crystallinity were computed and graphed below:

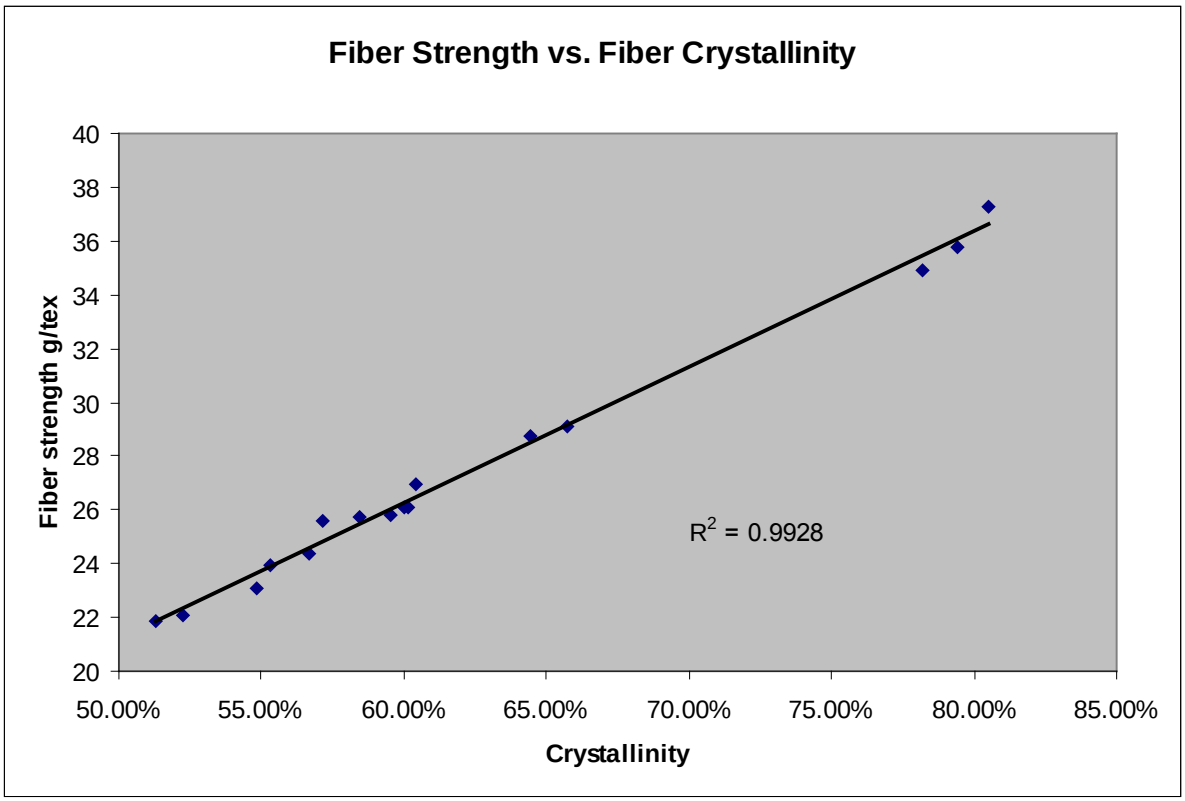


Figure 3.5. Fiber Strength vs. Fiber Crystallinity.

A linear trendline was chosen since it best demonstrated the relationship between strength and crystallinity. The graph revealed that the fiber strength increased with high percentage crystalline fibers.

The Egyptian fibers, in particular, show the greatest variation among the other fibers, seen as the three data points at the top right of the trendline.

3.7 Crystallite Size

Estimates of the crystallite dimensions normal to the *hkl* planes are based upon peak broadening which requires the measurement of the peak width at half maximum intensity[65]. Valid measurements of the crystallite sizes of cotton fibers should be made when overlapping peaks have been resolved and separated from the background scatter[66]. The observed widths of x-ray diffraction peaks are a result of two contributions, instrument broadening and broadening due to crystallite size. Reasons for instrument broadening include finite sample thickness, divergence of the incident beam, and the distribution of energy in the incident radiation[67]. As the size of the crystallites in the sample decreases, the x-ray diffraction peaks broaden. However, because of the contribution of instrument broadening (which should be relatively negligible), the crystallite size estimates in Table 3.6 should be considered as lower limits. An expression for determining the crystallite size from measurement of the FWHM was derived by Schrerrer:

$$\text{Crystallite Size} = \frac{k\,\lambda}{}$$

Using Scherrer's equation, the crystallite size of each cotton sample was observed to vary only between 35 Å and 40Å. The 002 peak was narrower in some cotton samples, giving rise to a bigger crystallite size estimate. No obvious correlation was observed between crystallite size and country of origin or fiber strength. There is no reason to believe that the crystallite size of the cotton fibers grown in various countries should be related to the orientation of microfibrils or the strength of the fiber.

3.8 Crystallite Orientation

It is generally accepted that the orientation of the crystallites with respect to the fiber axis is associated with the strength of the cotton fiber[68,69]. It is the most important structural parameter influencing mechanical properties. The molecular orientation in cotton is defined by the angle of the helix, which the crystallites make with respect to the fiber axis[70]. Methods used to determine crystallite orientation have been developed by Creely et al[71] and by Hermans[72]. Creely's method is based on the assumption that the distribution of the x-ray scattering intensity from the crystallites around the chi arc gives a measure of the orientation of the crystallites. Since there are a large number of (mostly) oriented fibers in the x-ray beam, the diffraction gives an average of the

orientation of the crystallites in the fibers. Thus the measurement of the FWHM of the 002 arc along the chi direction, called an azimuthal scan, satisfies the method.

However, the degree to which individual cotton fibers are alligned during sample preparation could also affect the observed crystallite oreintational distribution. Experimentally, the crystallite orientation (FWHM) relative to the fiber axis was observed to range from 42° to 80° (Table 3.8), indicating significant differences between the cottons. When plotted against fiber strength, the fiber orientation showed a relatively smooth, but non-linear correlation.

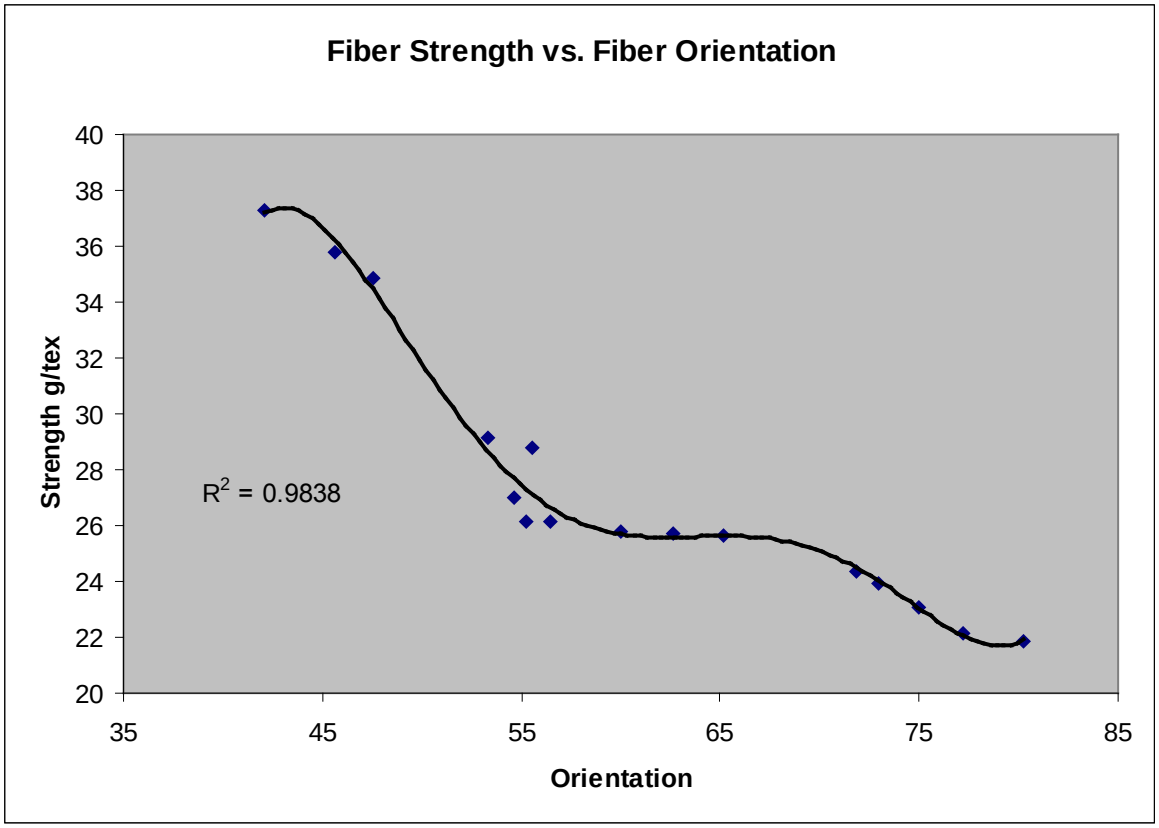


Figure 3.7. Fiber Strength vs. Fiber Orientation.

A polynomial trendline was chosen for analysis, as it best exposed the relationship between strength and crystallite orientation in cotton fibers.

The graph reveals that lower values of the orientation angle correspond to increased orientation of the cellulose crystallites to the fiber axis, and consequently, stronger fibers. The Egyptian cottons had the lowest orientation angles and the highest bundle tenacity.

3.9 Conclusion

With regard to fiber quality among all the countries observed, Egyptian cottons were far superior to all the other cottons. The use of synchrotron radiation afforded great precision during data collection in less time. Review of the plots indicated there were parameters that provide a clear distinction between Egyptian cottons and cottons from other countries. Fiber orientation and percent crystallinity proved to be the structural elements that provided the best information to distinguish Egyptian cottons from all other cottons tested. These structural differences were derived from the diffraction patterns measured for the cotton samples of each country. However, other than the Egyptian cottons, it would be difficult to take unknown cotton fibers from a country and try to guess what country it came from. In addition, the variation of fibers harvested in different seasons, in different soils and irrigation conditions, and grown in different regions of a country remains unexplored. However, the hypothesis that Egyptian cottons can be distinguished by x-ray fiber diffraction measurements appears to have been satisfied.

ACKNOLWEDGEMENTS

We would like to thank to Mr. Zimmerman and the U.S. Customs Agency for the financial support making the project possible. We appreciate the correspondence and technical help from Judith Brockmann, Randall Breaux and Charles McCombs from the U.S. Customs and Border Patrol Laboratory in New Orleans. Thanks also to the good people at the Southern Regional Research Center for providing direction to the relevant literature, answering all my questions, and providing the fiber strength data.

CHAPTER IV

DETERMINING THE CRYSTAL STRUCTURE OF CELLULOSE III_I BY MODELING

4.1 Introduction

Cellulose was the first carbohydrate to be studied by computer modeling. In 1960, Jones[73] used standard bond lengths, angles and interatomic distances to construct models that were used as part of a mostly unsuccessful attempt to solve the crystal structure of ramie cellulose I from fiber diffraction data. The advantages of the method were clear, however, and since then, computer models have been an integral part of most fiber diffraction studies that seek to determine the atomic positions.[74] Augmentation of crystal structure determinations by modeling is often necessary because the small number of diffraction intensities from most fibers is inadequate to determine the x, y and z coordinates of all unique atoms in the structure. With a combined approach, diffraction data can provide some guidance and the modeling energy calculations supply the rest of the information. This approach has been taken to the logical extreme of attempting to solve structures of small organic molecules by modeling with no specific experimental data whatsoever.[75] Those efforts are as yet not sufficiently reliable for general use but are at the forefront of modeling development.

As modeling has become more sophisticated, methods for experimental study of crystalline fibers have also improved. New sources of highly crystalline cellulose have

been identified, and the preparation of films of oriented crystallites allows the use of these crystallites regardless of their initial lack of orientation[76].

Neutron diffraction work has yielded the details of the hydrogen bond networks and very powerful synchrotron x-ray beams provide more diffraction data than laboratory generators. Together, the new techniques have resulted in sufficient data that high-resolution, model-free structure determinations of cellulose structures could, in principle, be carried out.

High-resolution structures are now available for cellulose Iα[77] and Iβ,[78] as well as cellulose II.[79] Most native cellulose is a mixture of the Iα and Iβ structures, with the Iα form being prevalent in cellulose that is produced by algae and bacteria, whereas Iβ is dominant in higher plants. The sample for the high-resolution study of cellulose II was produced by treating native cellulose I from flax with 23% NaOH, followed by rinsing and drying. Cellulose II can also be prepared by precipitation from solution, as in the manufacture of rayon, and by bacteria that are either mutants or at low temperature. A third major form, cellulose III, results from treatment with amines that are subsequently evaporated or rinsed off. Although their diffraction patterns are similar, subtle differences distinguish cellulose III that is made starting with cellulose I (III_I) from that starting with cellulose II (III_{II}). Finally, cellulose IV can be prepared by heating the other forms in glycerol at 260 C°. Recently, Wada et al. proposed that IV_I is actually Iβ with lateral disorder.[80]

In 2001, Wada et al. proposed that cellulose III_I has a single chain monoclinic unit cell with $P2_1$ symmetry and that the O-6 atoms were in the gt position.[81] Those results

contradict a 1976 determination by Sarko et al., who had done a complete analysis based on limited X-ray diffraction data.[82] Their work was based on a two-chain unit cell and determined the O-6 groups to be in tg orientations.

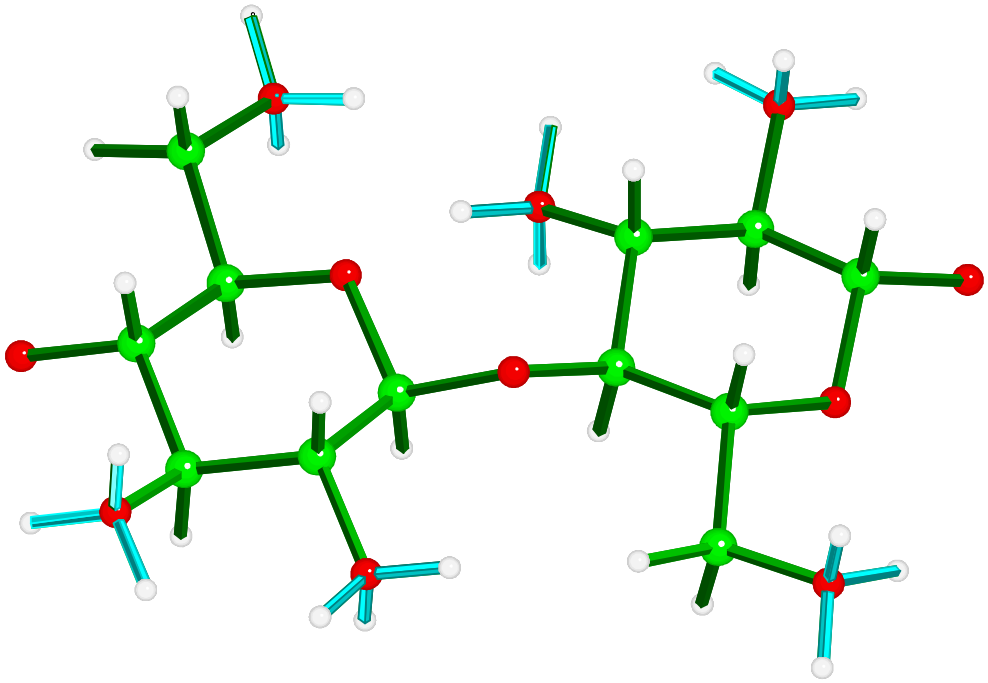
Although the pattern of Wada et al. has more than 100 intensities, they did not, in that work, attempt to solve the structure. Instead, the O-6 position was determined by accompanying NMR studies. Their results presented a unique opportunity. A modeling study could be independently carried out with an unknown that would inevitably be determined at high resolution. If successful, it was hoped that our project would encourage the incorporation of higher-quality modeling methods in fiber diffraction studies. These combined methods would continue to be of use on less-crystalline samples. Of course, a successful prediction would lend credibility to modeling studies on other materials such as amorphous cellulose, for which experimental data are limited and more difficult to interpret.

The high-resolution experimental study of cellulose III has now been published,[83] and we can also compare those results with ours, which were presented at two meetings.[84]

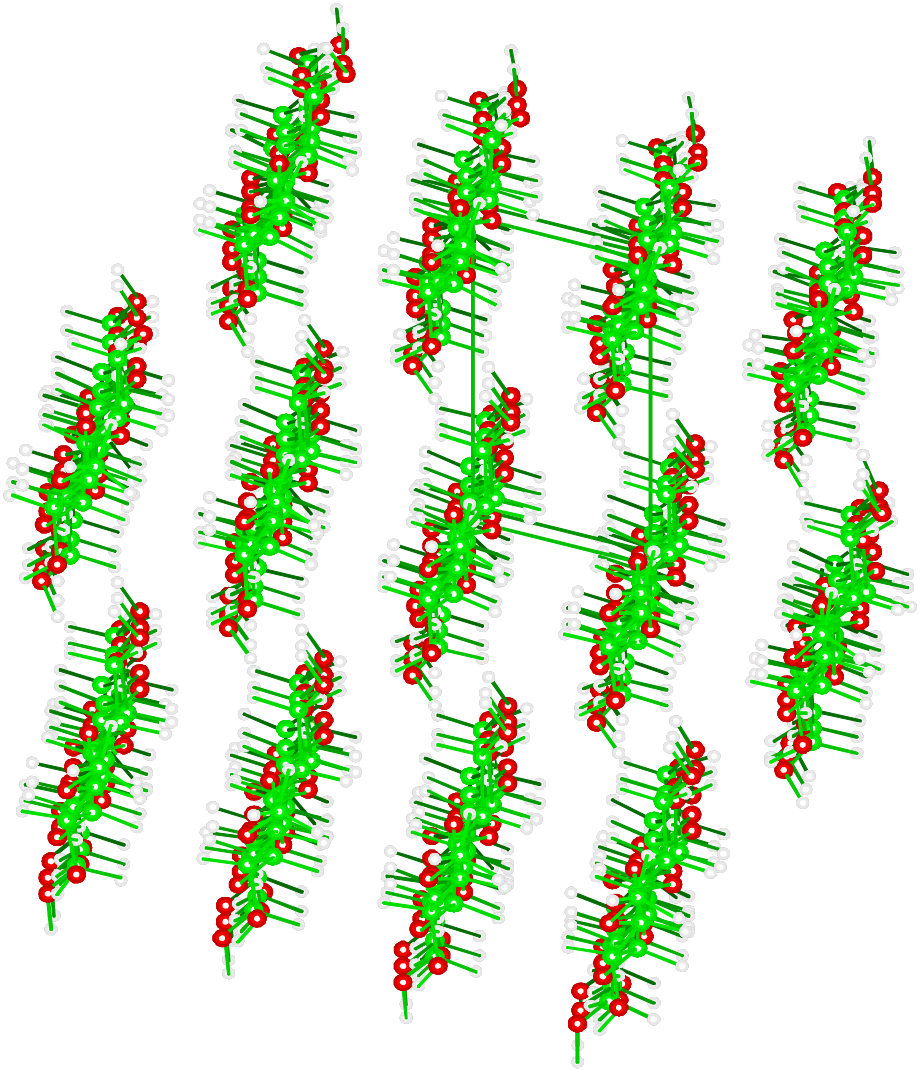
4.2 Methods

Given the results from Wada et al. regarding the O-6 position and unit cell dimensions and symmetry,[81] only the hydroxyl group orientations remain as explicit variables. Cellotetraose molecules were constructed with Chem-X with two-fold screw-axis symmetry and capped with methyl groups at the reducing and non-reducing ends

to prevent the formation of unrealistic hydrogen bonds. The O-2, O-3 and O-6 hydroxyl groups on the tetraose models were placed in each of the three staggered orientations (Figure 1), so that they made torsion angles of –60°, 60° and 180° with the H2, H3 and C5 atoms. Thus, there were 27 combinations of hydroxyl orientations.



These models were placed visually in the unit cell according to Figure 4.5 in Wada et al., in both the “up” and “down” orientations,[85,86] for a total of 54 starting models. There was substantial confidence in the orientation presented by Wada since it was based on the report by Sarko et al.[82] That orientation would depend on the very strong hk0 reflections and is likely to be unaffected by other errors in the determination. Symmetry operators within Chem-X were used to generate clusters (minicrystals) with 13 chains, similar to previous designs[87], as shown in Figure 4.2. These 54 minicrystals were then each energy minimized with MM3(96), using a dielectric constant of 3.5 and the hydrogen bonding potential from MM3(92). We have found that those modifications result in better model crystal structures. No constraints, symmetry operators or periodic boundaries were placed on the structure during minimization. The plan was to observe the resulting energies and hydrogen bonding schemes and to select one or more likely structures for comparison with the proposed two-chain structure from Sarko et al.



The minicrystal method is subject to uncontrolled edge effects[88] regarding the positions of the external atoms. However, it has the advantage that it can provide energies that are based on a variety of different potential energy functions, including MM3, which is known to reproduce a number of phenomena related to carbohydrates. All energies are reported as kcal/mol of the structures in question. Thus, the energies reported for the tetraose-based minicrystals would be kcal for a mole of minicrystals. Other energies reported include kcal/mol of hexaose-based minicrystals and kcal/mol of a layer of cellobiose residues inside the hexaose-based minicrystal. These energies are reported below simply as kcal.

4.3 Results and Discussion

Of the 54 models based on single-chain unit cells, 16 gave total minimized steric energies that were between 237 and 246 kcal. Eight of these were up models, and the other eight were down. A second group of 26 had energies between 318 kcal and 367 kcal, and the remaining structures had energies between 407 and 470 kcal. Only the group with energies of about 240 kcal is relatively homogeneous in energy and hydroxyl orientation. That homogeneity is an additional confirmation that the lowest energy group represents the most likely structures. Table 4.1 shows that the best up model has an energy of 237.6 kcal, whereas the best down model has an energy of 236.7 kcal. These

values can be compared to the energy of the minimized, tetramer-based model of Sarko et al., 340.3 kcal.

Table 4.1. Energies (kcal) and hydroxyl torsion angles (º) for two central glucose residues from the best tetraose-based models

Model	Energy kcal	τ_2	τ_3	τ_6	$\tau_{2'}$	$\tau_{3'}$	$\tau_{6'}$

orientation. For example, the hydroxyl groups on C2 and C2' rotated from initial values of 60° to final values near to 12°, a rotation of 48°.

Hydroxyl groups on C2 of other structures in the low-energy group rotated to the same values near 12° starting from -60°, a rotation of 72°. The corresponding rotations at C3 and C6 of the lowest-energy structure were more than 107° and more than 40°, respectively. Hydroxyls on C6 atoms in other structures started at -60° and rotated about 72°. The extents of rotation of the hydroxyl groups were surprising since they were initially in staggered positions, normally considered to be energy minima, although nearly eclipsed conformations, such as the 12° torsion for O2H, are fairly common in carbohydrates and cyclitols.

Such large rotations during minimization indicate that the attractiveness of the hydrogen bond system was so great that the hydroxyl groups overcame energy barriers. The similarity of the unprimed and primed torsion angles in Table 4.1 strongly supports the experimentally determined two-fold screw-axis symmetry.

Unit cell dimensions were assessed based on the interchain distances and angles. Those that were based on tetramer models were approximately a=4.5$\pm$0.09, b=8.0$\pm$0.1, c=10.35$\pm$0.03, α=90.1$\pm$2, β=90.0$\pm$1.0, and γ=105.5$\pm$0.4 for the minimized models. Comparisons with the experimental values listed in Table 4.2 were satisfactory. Our minimized version of the model of Sarko et al.[82] gave a=10.44, b=7.95, c=10.36, α=90.3, β=89.8, γ=122.85. Differences from the experimental values in Table 4.2 were

also considered minor. The slight expansion of the unit cells, particularly along the a-axis, may be partly due to the lack of long-range packing forces in the minicrystals.

Table 4.2. Calculated energies and unit cell dimensions of hexamer models.

Hexamer Model	Minicrystal Energy (kcal)	Cellobiose Layer Energy (kcal)	a(Å)	b(Å)	c(Å)	α(°)	β(°)	γ(°)

Because of the shifting in the two-chain cell, its chain ends would not experience the same degree of stabilization from van der Waals attraction to their neighbors, as would the coplanar ends in the one-chain cell models. That problem was solved by comparing the energies of internal cellobiose layers in minicrystals that were built from methylated cellohexaose molecules.

The energies for the cellobiose layer were based on subtraction of the energies of the best up and down methylated cellotetraose minicrystals from energies from analogous methylated cellohexaose minicrystals. Those cellobiose layer energies, which do not have first-order end effects, are shown for the one- and two-chain cell structures in Table 4.2, along with the unit cell dimensions of the models based on the cellohexaose molecules. In this case, the energies of the “up” structure, both the full hexameric minicrystal and the cellobiose layer in the minicrystal, were slightly lower than those of its “down” counterpart but considerably lower than those of the two-chain cell structures.

Table 4.3 shows the geometries of the hydrogen bonds in which the central cellobiose unit in the minicrystal is involved, based on the hexameric models. Based on the criterion that the distance between the donated hydrogen and the acceptor oxygen atom is < 3.0 Å and the O—H…O angle is > 90°, there are three intramolecular and two intermolecular hydrogen bonds.

Table 4.3. Intra - and Intermolecular[a] hydrogen bonds in best “up” model.

Type of bond	H-Bond	Length H…O (Å)	Length O…O (Å)	Angle (°)

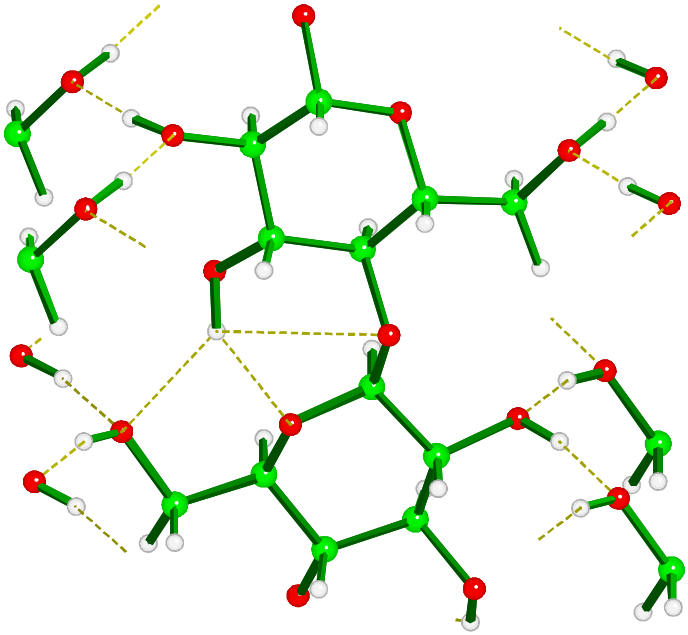
That frequently overlooked interaction stabilizes the gt position in many related molecules, despite H…O distances that are longer than are often considered to be hydrogen bonds.[90] The third intramolecular interaction in Table 3, O3-H…O-4 is indeed very weak, but its presence is noted.

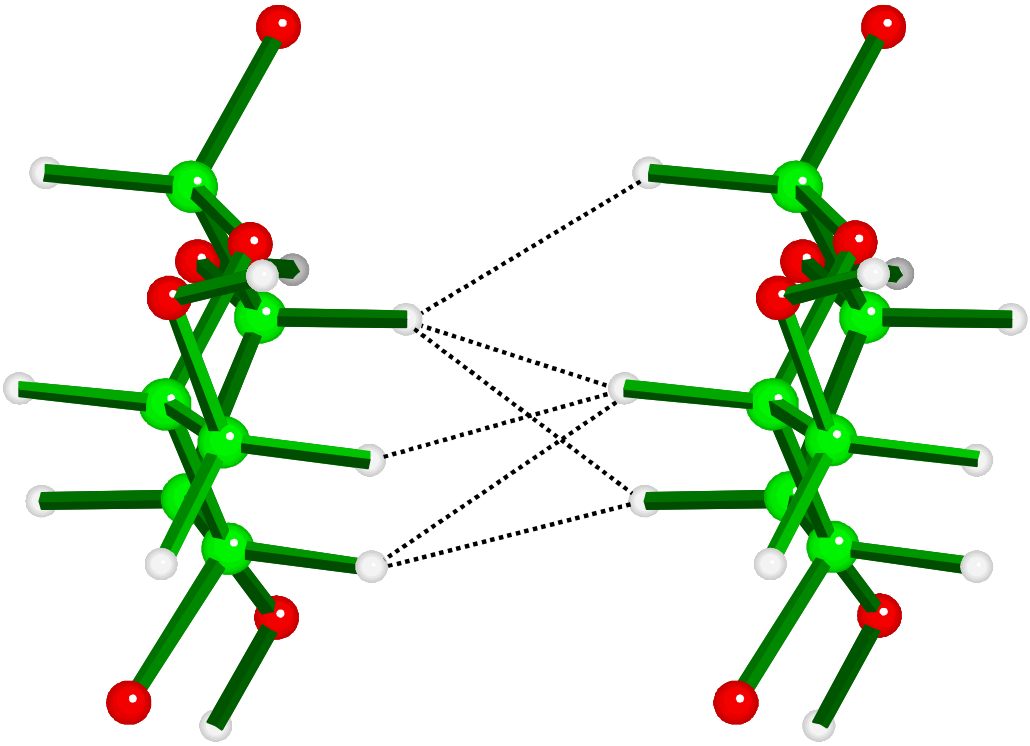
Although Table 4.3 shows four intermolecular hydrogen bonds in which the central cellobiose is the donor and four hydrogen bonds in which it is the acceptor, there is just one unique O6-H…O-2 hydrogen bond and one O2-H…O-6 hydrogen bond when there is actual two-fold symmetry. The near identity of these modeled geometries for the O6-H…O-2 hydrogen bonds confirms that the two-fold, single chain structure is consistent with the MM3 force field. The O2-H…O-6 geometries lead to a similar conclusion.

The intermolecular hydrogen bonds participate in “infinite” chains of donor-acceptor-donor linkages (Figure 4.3) that have excellent hydrogen bonding geometry. Such systems have increased strength and shortened interatomic distances because of the phenomenon of “cooperativity”.[91]

Van der Waals forces are also important, with stacking of the residues in the a-axis direction. Each of the methine hydrogen atoms is in van der Waals contact with one or more methane hydrogen atoms on the neighboring molecules. Figure 4.4 illustrates the H…H distances > 3.2 Å for the best up model.

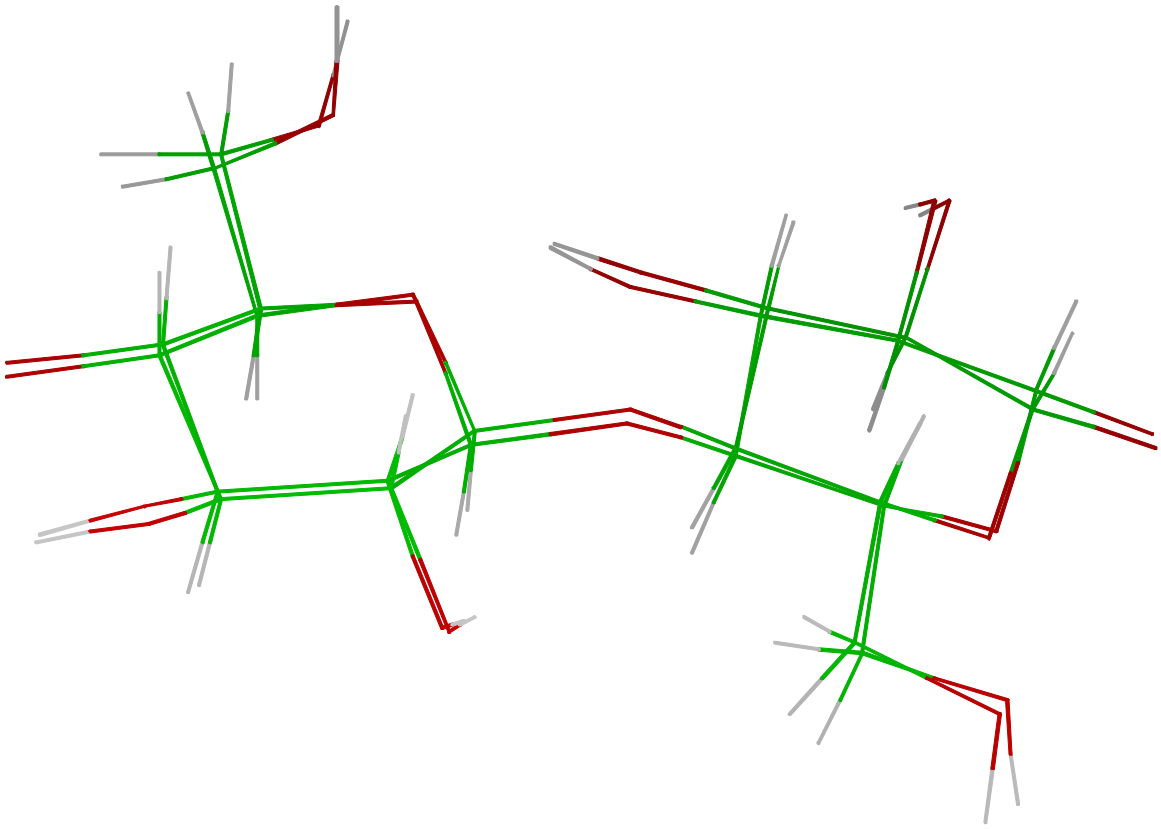
Our best “up” model is similar in many respects to the high-resolution structure very recently published by Wada et al.[83] Interestingly, they were able to clearly rule out the down packing model, while our results were ambiguous on that point.





The conformations of the primary alcohol groups in the experiment and model were 44° and 59°, respectively. Despite that difference, the resulting positions of the O-6 hydroxyl hydrogens are quite similar. The biggest difference is in the positions of the two protons attached to C-6. These relationships are shown in Figure 4.5, in which the central cellobiose unit from the hexaose-based up minicrystal is fitted to a cellobiose unit generated from the coordinates of Wada et al.[83] The root mean square difference between the positions of the 12 ring atoms and the linkage oxygen is only 0.1 Å.

In the high-resolution structure of Wada et al.,[83] there was one slight ambiguity regarding the direction of the infinite cooperative hydrogen bonding network. Although their final result was quite similar to ours, they also considered an alternative that reversed the direction of the donor and acceptor hydroxyl groups. In the agreed upon network, our O-2 hydroxyls have 12° torsion angles, nearly eclipsing the C2-H hydrogen atoms. In the alternative network structure, the O2-H atoms are oriented anti to the C2-H hydrogens. Experimentally, this ambiguity arises because of the difficulty in precisely locating the proton between two oxygen atoms. If it is closer to O-2, then it is taken to be covalently bonded to O-2 and hydrogen-bonded to O-6, and vice versa. In a structural or modeling sense, direction of the hydrogen bonding in an infinite network is expressed by the rotational orientation of the hydroxyl groups. The modeling results were less ambiguous, because the various torsional and other steric terms in the force field resulted in the alternative systems being considerably higher in energy. Several minicrystals having the alternative hydrogen-bonding scheme fell into the second lowest-energy group.



To understand why Sarko et al. proposed a two-chain structure, we reviewed their published (as supplementary data) crystallographic information and recorded a fiber diffraction pattern of ramie cellulose III prepared by the method of Calamari et al.[92] All of the diffraction spots on our low-resolution pattern could be indexed with the one-chain cell. All but two of their first-layer line spots (d-spacings = 2.78 Å and 2.55 Å could also be indexed with the one-chain cell. Those spots were not visible on our pattern. Their published pattern does not permit a close analysis, but one plausible explanation, that traces of cellulose I remained, is not likely because there is no 2.78 Å observed hk1 spacing from cellulose I.[93] It appears that Sarko et al. assumed that there were two chains in the cell. Ironically, Sarko and Muggli had earlier discussed a one-chain unit cell for cellulose I before the distinction between cellulose Iα and Iβ was understood.[94] In any case, the synchrotron fiber diffraction pattern by Wada et al. produced 114 reflections that were indexed by the proposed one-chain monoclinic unit cell, compared to Sarko's 23 reflections. The cell based on the larger number of reflections should overrule one based on so many fewer spots.

Our molecular modeling study of cellulose III_I concurs that the unit cell of Wada et al. is the more probable. However, our best up and down models show very small differences between them, either in the energies, the unit cell values, or the hydrogen bond geometries. Therefore, either model could correspond to the structure of cellulose III_I. The final modeled coordinates are listed in Table 4.4.

Table 4.4. Atomic coordinates of modeled glucose monomer of up cellulose III_I.

ATOM	X	Y	Z

HO2	1.52256	3.059395	3.31848

Table 4.5. Atomic coordinates of modeled glucose monomer of down cellulose III_I.

ATOM	X	Y	Z

HO3	-1.71577	-3.78143	1.86147

BIBLIOGRAPHY