**Crystal Molecular Dynamics**

**by**

**Paweł A. Janowski**

**A dissertation submitted to the**

**Graduate School of New Brunswick**

**Rutgers, the State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Graduate Program in Chemistry and Chemical Biology**

**and**

**Graduate Program in Computational Biology and Molecular Biophysics**

**Written under the direction of**

**Prof. David A. Case**

**and**

**Prof. Darrin M. York**

**and approved by**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**New Brunswick, New Jersey**

**August, 2015**

# Abstract of the dissertation

**Improved Molecular Dynamics and Macromolecular Crystallography through Simulations of Biomolecular Crystals**

**By Paweł A. Janowski**

**Dissertation Directors**

**David A. Case Ph.D and Darrin M. York Ph.D**

We present a broad effort at the development of crystal simulation methodology and its application to benefit both macromolecular crystallography and molecular dynamics methods. Crystallography is the current method of choice for structural determination of biomolecules, but it is hampered by the inherently time and space averaged nature of the experiment as well as methodological limitations that do not sufficiently account for the heterogeneous and dynamic nature of crystals. Molecular dynamics has proven itself as a method capable of probing the physics and chemistry of biomolecules on an atomic scale, but requires continued development of the underlying force field parameters to more accurately reproduce observables. Our effort has focused on developing the framework for molecular dynamics simulations of biomolecular crystals. We first present our methodology for performing crystal simulations and show how it is applied first to simple peptide crystals and then to increasingly complex biomolecular systems. Next we demonstrate the utility of crystal simulations for validation of molecular dynamics methods through two case studies of the biophysics of enzyme reactions. Finally we demonstrate the improvement to crystallographic methods that can be gained by incorporating molecular dynamics methods. Our work is of great benefit to both the molecular dynamics and macromolecular crystallography communities and proposes specific approaches to integrate the two fields for the benefit of both.

# Acknowledgements

I thank the Lord for the beautiful gift of His Creation, for the flowers, the mountains, the creatures and the fascinating molecular mechanisms of life that testify to His love. Thank you for the gift of reason, free will and desire of the Good, the True and the Beautiful that drives us in all our pursuits. May we one day attain the fullness of Love.

I thank Maria my wife who accompanied me throughout this path in person and in spirit. Thank you for your care, your patience, your inspiration. We have truly walked the past five years together in friendship. May we walk many more!

I thank my parents, Jolanta and Andrew, for their love in raising me to be the person I am. Thank you for teaching me what is important in life and what is not. I am forever indebted in the bond of filial love.

I thank my family and my friends for all your kindess, your love, all the good times spent together that will forever form part of the treasure of my memories. Without those wonderful times shared together, I would have been hard pressed to keep my sanity along the way.

I thank all my colleagues, my lab mates, my teachers, all the wonderful scientists I have had the pleasure to work with. Thank you for showing me how exciting science is!

I thank all of the good people I have met. Each encounter with each one of you enriches both of us. Only in relationship with others does man become truly man. Never stop being who you are. I love you all.

Finally, and in a special way, I thank my advisors, Prof. David Case and Prof. Darrin York. You have challenged and inspired me constantly and at the same time been caring guides along the path. I have learned so much from you, not just about science but about what it means to be good human beings. I will remain forever grateful.

To the loving memory of my Father who would have wanted to be here but is even closer than we can imagine.

Contents

[Abstract of the dissertation ii](#_Toc423611938)

[Acknowledgements iv](#_Toc423611939)

[Contents vi](#_Toc423611940)

[Abbreviations used vii](#_Toc423611941)

[Section I. Introduction 1](#_Toc423611942)

[Chapter 1. Introduction and background 1](#_Toc423611943)

[Section II. Developing molecular dynamics of crystals. 8](#_Toc423611944)

[Chapter 2. Fav8 1 8](#_Toc423611945)

[Chapter 3. Fav8 2 8](#_Toc423611946)

[Chapter 4. 4lzt 8](#_Toc423611947)

[Chapter 5. DNA/RNA 8](#_Toc423611948)

[Section III. Applications of molecular dynamics of crystals 8](#_Toc423611949)

[Chapter 6. Hairpin 8](#_Toc423611950)

[Chapter 7. RnaseA 8](#_Toc423611951)

[Section IV. Improved crystallographic methods through crystal molecular dynamics 8](#_Toc423611952)

[Chapter 8. AFITT 8](#_Toc423611953)

[Chapter 9. Phenix-Amber 9](#_Toc423611954)

# Abbreviations used

MD – Molecular dynamics;

BX – Biomolecular crystallography;

FT – Fourier transform;

Å – Ångstrom;

# Introduction

## Introduction and background

When, during my initial visit to Rutgers, Prof. David Case first mentioned the idea of improving crystallography through molecular dynamics of crystals, I felt a tinge of excitement. I had studied crystallography for two semesters during my undergraduate coursework at Jagiellonian University in Krakow. Lectures were eloquently delivered by one of the best teachers I’ve ever had, Prof. Krzysztof Lewiński. However, despite all my effort I could not grasp the essence of how a seemingly random pattern of dots on a sheet of paper could be turned into a three dimensional model of a biomolecule. I liked crystallography, but I also respected it and I feared it because I felt like there was something powerfully beautiful and mysterious about it. So when Dr. Case floated this idea of molecular dynamics of crystals I was excited: here was a chance to make up for my previous failing, to finally come to understand crystallography or to die trying. And to do that by using the molecular dynamics that I wanted to focus my Ph.D. studies on… it was the perfect project.

Thus I have happily spent the last five years focused on our effort to simulate biomolecular crystals with molecular dynamics. The original question we asked ourselves was simple: what can we learn from molecular dynamics of crystals? And this was quickly reformulated into the following four overarching questions that form the focus of this work:

1. What is the best way to carry out molecular dynamics of biomolecular crystals?
2. How can we use crystal simulations to improve molecular dynamics methods?
3. How can we use crystal simulations to improve crystallography methods?
4. What can we learn about real crystals from our simulations of crystals?

What follows is a brief introduction to the methods of crystallography and molecular dynamics, with special emphasis on aspects that relate directly to our work. We then discuss the goals and specific aims of this research and present the general organization of the dissertation before moving on to a presentation of the work in subsequent chapters.

### Crystallography background

Crystallography is a biophysical technique used to probe the three-dimensional distribution of atoms in molecules by analyzing the diffraction pattern of electromagnetic radiation on a crystal. As the name implies, crystallography requires that billions of copies of the molecule of study arrange themselves in a regular repeating array which is, by definition, a crystal. When used to study the structure of biomolecules, the method is referred to as macromolecular crystallography (MX). The fact that protein molecules can form crystals has been known for almost 150 years. In general, crystal formation of biomolecules is promoted by slowly removing solvent from a solution of the protein of study. If the solvent is removed too quickly or if the solution is not of the required purity, the protein molecules will precipitate out of the solution and form an amorphous powder. However if the solution becomes supersaturated slowly the molecules may pack themselves in a regularly repeating array held together by non-covalent chemical interactions in a way that minimizes the overall energy of the solute. Finding the exact conditions under which a given biomolecule crystallizes can be very challenging and in many cases constitutes the crux of the crystallographic method.

Once crystallized, the regularly repeating array of the crystal acts as a diffraction grating when light is shined upon it. Diffraction refers in general to the physical behavior of waves as they impact objects or slits. Etimologically, the term was coined by Francesco Maria Grimialdi in 1660 and comes from the Latin diffringere meaning “to break up into pieces”. In particular a regularly spaced array of slits or objects will cause the waves scattered off each object to interfere with each other. Wave crests lining up leads to constructive interference resulting in waves of higher amplitude, whereas when crests and troughs mix, destructive interference results in low amplitudes. Because of the dual nature of electromagnetic radiation, when light shines on diffraction grating it behaves like a wave and interference leads to the formation of bands (in the case of a one-dimensional diffraction grating) or spots (in the case of a two-dimensional diffraction grating). James Gregory’s observation of the diffraction pattern of light shining through a bird feather in the late 17th century constituted the discovery of the first diffraction grating.

A crystal is a repeating array of objects and thus can naturally act as a diffraction grating. However, because the wavelength of visible light is much larger than the typical spacing between array planes in molecular crystals, the diffraction of light on molecular crystals is not observed. The breakthrough moment for crystallography came in 1912 during a conversation between Paul Peter Ewald and Max van Laue, when van Laue suggested that x-rays (discovered in 1895 by Wilhelm Roentgen) might have a shorter wavelength that would allow their diffraction on crystals to be observed. In 1912 van Laue recorded the first ever x-ray diffraction pattern on a copper sulfate crystal. Shortly thereafter the father-son pair of William Lawrence Bragg and William Henry Bragg formulated the law that describes the diffraction of x-rays on a crystal. The first diffraction pattern from a protein crystal was obtained by John Desmond Bernal and Dorothy Hodgkin using pepsin, and the first three-dimensional structure of a protein molecule solved using x-ray crystallography was myoglobin in 1958 by John Kendrew. Van Laue, the Braggs and Kendrew all received Nobel Prizes for their work. In all thirteen Nobel Prizes have been award for work on or using crystallography.

The raw experimental data obtained in a crystallography experiment is a diffraction pattern (fig 1). This pattern is obtained as a beam of x-rays is focused on a crystal and the x-ray photos scatter (diffract) off the electron clouds of the atoms that make up the crystal. For the work presented here it is crucial to understand that the diffraction pattern is not obtained in a single instant from single x-rays scattering off the crystal. Rather it is obtained over a significant period of time usually ranging from a few up to about 30 minutes. The diffraction spots themselves require the constructive interference of a enormous number of x-rays to be observed. Furthermore the xrays themselves diffract off the billions plus molecules that make up the crystal. Thus one can say that crystallography is truly a time and space averaged experimient.

The diffraction patterned obtained in the crystallography experiment contains to essential pieces of information. The first of these is the location and spacing of the diffraction spots. The spots appear on the vertices of an array called the reciprocal space lattice which is a mathematical construct directly related to the parameters of the real space lattice. The real space lattice is the lattice of the crystal being studied. The appearance of diffraction spots can be described via the equation that is know as Bragg’s Law (named for the younger of the two Braggs mentioned above). Essentially, diffraction spots can only form in locations where the x-rays arrive in phase (in more simple language, where the crests and troughs of the arriving x-ray waves are lined up with each other). The condition for this to happen is that the distances of the paths that all the arriving x-rays travelled must all differ by an integral number of wavelengths (wavelength is the distance from one crest of the wave to the next) of the x-rays. This is presented schematically in Figure xx. The resulting description of the formation of diffraction spots is Bragg’s Law:

|  |  |  |
| --- | --- | --- |
|  |  | (Eq. ) |
|  |  |  |

where d is the spacing between a given set of planes in the array, θ is the angle at which the x-rays impact the set of planes, *n* is a positive integer and λ is the wavelength of the x-rays. The lattice (spacing between planes) and wavelength are constant under normal experimental conditions. Thus they uniquely specify the angle at which the scattered x-rays interact constructively and form a diffraction spot. Each spot thus corresponds uniquely to a specific set of planes in the array. Furthermore the angle is inversely proportional to the spacing. In other words smaller diffraction angles correspond to larger plane spacing in the lattice. Diffraction spots closer to the center of the diffraction pattern carry information about larger-scale features of the crystal. This is the basis for the concept of resolution: usually the diffraction pattern is only measured up to a certain radius: beyond that the angle of diffraction is too large and the spots too weak to be reliably recorded. Most importantly, by accurately measuring the location and spacing of the diffraction spots, one can deduce the spacing of the crystal’s array and thus obtain the parameters of the crystal unit cell (the three box dimensions *a*, *b*, *c* and three box angles α,β,γ.

The other essential information in the diffraction pattern are the intensities of the diffraction spots. Where the location of the spots reveals the unit cell parameters of the crystal array, the intensities of the spots tell us about the actual distribution of scattering objects, i.e. atoms, within each unit cell. The intensity of wave is equal to the square of it’s amplitude:

|  |  |  |
| --- | --- | --- |
|  |  | (Eq. ) |
|  |  |  |

We know from the previous discussion and Bragg’s Law that an identical scattering object located at each lattice plane of a certain spacing *d* would produce an ideal constructive interference between x-rays and consequently a diffraction spot at angle θ. But what happens if there are additional scattering objects located between the planes (Fig 1). The x-rays scattering off these objects will arrive at the diffraction location with a phase different from that of the rays scattering from the primary object. The resulting amplitude of the x-ray wave arriving at the diffraction spot location is obtained by summing the waves diffracted of each object within the crystal unit cell. Because the objects do not all lie integral distances of the scattering plane away from each other, the resulting waves that are summed are not all perfectly in phase. This results in an attenuation of the amplitude of the resulting wave and in extreme cases (Fig…) can results in a complete disappearance of the spot. If we treat the scattering electron density in the unit cell as continuous and divide it into infinitesimal sections *dx* along the scattering vector, the amplitude of the resulting diffraction spot can be obtained by integrating the partial x-ray wave scattered by each section *dx* of electron density:

|  |  |  |
| --- | --- | --- |
|  |  | (Eq. ) |
|  |  |  |

(Eq. 3) is presented for the one dimensional case but the generalization to three dimensions is straightforward. Each partial wave has an amplitude proportional to the electron density at *x* but with a phase relative to *x=0* of *2πhx*. The integration is performed over the unit cell vector *h* and the position *x* are described in fractional coordinates. ρ(x) is the electron density at position *x.* ***Fh*** is called the structure factor and is a wave described by an amplitude  and a phase . The intensity of the diffraction spot is related to the structure factor amplitude via (Eq. 2) As it turns out, this equation is equal to the mathematical transformation known as the Fourier Transform (FT). Conversely if we sum over each one of the diffracted waves (at each diffraction spot), we obtained the scattering electron density:

|  |  |  |
| --- | --- | --- |
|  |  | (Eq. ) |
|  |  |  |

Here again we present the one-dimensional form for pedagogical purposes. The summation is over all the diffraction spots of order *h*. This equation corresponds to the form of the inverse Fourier Transform and is the mathematical inverse of the (Eq. 3). Thus we arrive at one of the fundamental concepts of x-ray crystallography: the electron density of the crystal unit cell is the inverse FT of the diffraction pattern.

Now let us examine what is needed to calculate the electron density. (Eq. 4) states that we need to perform a summation over each diffraction spot. For each spot we need the amplitude and phase of its corresponding structure factor. The amplitude is readily obtained as the square root of the intensity measured in the experiment, but unfortunately there is no information about the phase. This is known as a phase problem. Many ingenious (and difficult to implement) methods exist to tackle the phase problem. Here let it suffice to say that if a sufficiently good estimate of the phases is obtained from which a sufficiently good estimate of the electron density can be calculated, then one can move on to the next part of the process, refinement, that is of much greater concern to us in the present work. In practice, the great majority of biomolecular structures are solved today by a technique called molecular replacement where a sufficiently good initial estimate of the electron density and phases is obtained by comparison to another similar molecule whose structure is already known.

Let us know suppose that we have a prOnce an estimate model is obtained, the next stage is refinement which is most pertinent here. How can we continue once we have that? FT forward need electron density at every point on a 3D grid. This can be calculated from the distribution of atoms using the …. FT of ED gives us the amplitudes and phases. We don’t know about the phases. But we know the amplitudes. We can compare the two. Usually for this comparison a statistic called the R-factor is used. Smaller the R, the smaller the sum of differences between the observed and calculated amplitudes (Fobs, Fcalc). If not perfect, we can move the atoms around and recalculate the R-factor and keep moving until better.

We could do this by hand and it might work or we might be at it forever. Mathematical schemes to minimize. In the basic approach minimize the sum of square difference between the amplitudes (F(xyz)). Not well defined… parameter to observed ratio Alternatively we can use a maximum likelihood formulation. Advantage of allowing a Bayesian treatment. Minus log and minizimize is same as finding the max. The now becomes. Why prior? Because increases the observables… adds additional constraints, lowers the search space of the optimization algorithm.

Prior… most programs use EH. What is EH… But originally MD… Moved away from… (Maybe sec 3).

In practice more complicated. XYZ refinement not robust enough to find the optimal location- stuck in local minima. Rounds of refinement, manual rebuilding. Second many more parameters. Fluctuations modelled as B-factor (formula)… refine B-factors as isotropic or anisotropic. TLS parameters. Occupancy and alternate conformations. Bulk solvent and anisotropic scaling factor. Macrocycles… in Phenix… In any case one ends up with a 3D model of the locations of the atoms in the crystal.

### Molecular Dynamics Background

Molecular dynamics is... F=ma, U=… etc. Standard force field used in Amber… Each term means…

Additional stuff, such as thermostat, shake, etc.

Standard simulation is done in solvated box, not crystal. Usually insert, remove, equilibrate…

MD has shown itself to be remarkably useful and successful. In the core just a simplified model. So much not modelled, but … examples….

### Goals and overview

Both methods exteremely valuable but also suffer from limiations. These limitations can be overcome at least in part through MD of crystals. Let’s look at each one.

Crystallography: First sources of error and noise. Sometimes so high (low res) that indeterminate. Simulations if reliable could tell us more to help resolve. Second, time and space average… End up with single static view. Myopic because one best rep view of the average. In fact crystals move (dynamic) and heterogenous. Recent efforts by several group aimed at resolving this. Ensemble refinement. Networks stuff from Fraser. Diffuse scattering from … Move and insights about functions. Also insights about crytals. By simulating over multiple copies (space) and ns (time) we can undo the averaging and get time-space resolved glimpse. A more comprehensive view of the crystal. Also, information about crystals: solvent distribution, etc lead to better refinement techniques. Finally, in itself a better set of priors.

MD is good but usually run as solvated box. Good sense but drawback that not comparing against experimental… Validation from crystals. Look directly at experimentals such as structure, fluctuations, electron density and amplitudes… Can check how well it’s doing and modify ff…

Organization: Part II developing methodology for simulating crystals. Apply to larger and conclusions about MD… Part III crystal syms applied to scientific investigation : two cases where used to validate Part III md of crysals applied to improve cyrstallograpjhy methods… Copyrights…

# Developing molecular dynamics of crystals.

## Fav8 1

Blah blah

## Fav8 2

Blah blah

## 4lzt

Blah blah

## DNA/RNA

Blah blah

# Applications of molecular dynamics of crystals

## Hairpin

Blah blah

## RnaseA

Blah blah

# Improved crystallographic methods through crystal molecular dynamics

## AFITT

Blah blah

## Phenix-Amber

Blah blah