

Modeling Metal Protein Complexes from Experimental Extended X-ray Absorption Fine-Structure using Computational Intelligence

Collin Price

Department of Computer Science

Submitted in partial fulfillment
of the requirements for the degree of

Master of Science

Faculty of Mathematics and Science, Brock University
St. Catharines, Ontario

©Collin Price, 2014

Chapter 1

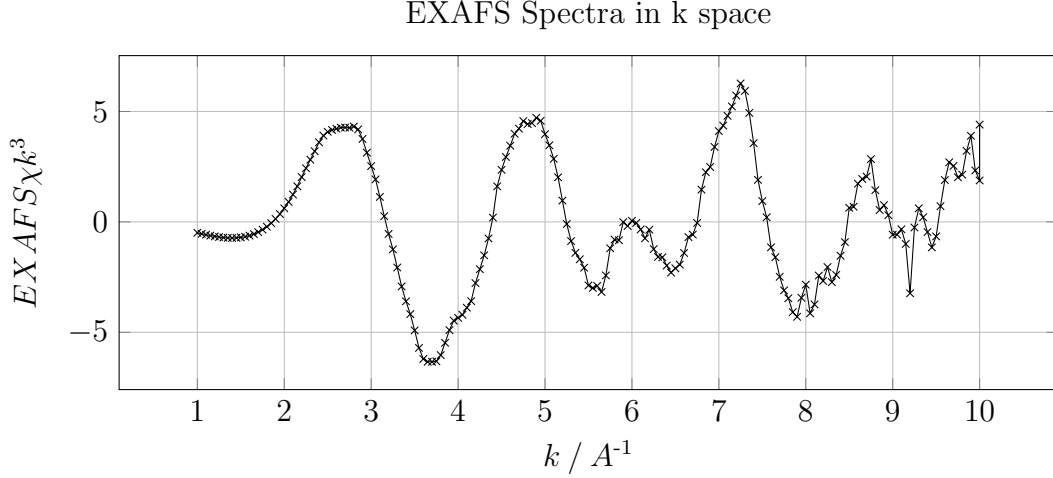
Introduction

The aim of this thesis is to find a better method for determining the atomic structure of a molecule using Extended X-Ray Absorption Fine Structure (EXAFS). The following thesis uses the oxygen-evolving complex (OEC) in state S_1 as an example for structure refinement but the developed process can be applied to any given structure that has an EXAFS spectra. In this chapter, we introduce the biological background and terms, followed by the problem definition, and finally elaborate on the computer science theories applied to the problem.

1.1 Biological Background

The oxygen-evolving complex (OEC) is the water-oxidizing enzyme of photosystem II. The responsibility of the OEC is to accept water molecules as input and output the oxygen and hydrogen atoms to be used in the system [1]. The incoming water molecules are put through four different states in order to perform the oxidization. The atomic structure of the OEC molecule is different in each state because each state has a unique role in the process.

(Work biomimetic chemistry into this. Define: mimicking the laws of nature to synthesize compounds.)

Figure 1.1: EXAFS Spectra of OEC in S_1

1.2 X-ray Absorption Spectroscopy

The following overview is based on information contained in Matthew Newvilles Fundamentals of XAFS (2004) [2]. X-Ray absorption fine structure (XAFS) is a method used to measure the absorption coefficient of a material as a function of energy. X-rays are part of the electromagnetic spectrum with wavelengths ranging from 25\AA to 0.25\AA . All atoms resonate at a specific wavelength. The x-ray is tuned to have the same wavelength as the target atom. A photon from an x-ray is absorbed by an electron in a tightly bound quantum core level of an atom. Absorption only takes place if the binding energy of the core level is less than the energy of the x-ray photon. At the time of absorption a core electron moves to an empty outer shell and another electron moves in to take its place. Eventually the affected electrons decay to their original state. During this time fluorescence energies are emitted that characterize a specific atom.

The absorption coefficients measured after the initial absorption are referred to as the EXAFS. During the decay of the electrons to their original state, oscillations occur in the measure of the absorption coefficient. The different frequencies found within the oscillations correspond to different near-neighbour coordination shells, which can be described and modeled according to the EXAFS equation. From the oscillations, the number of neighbouring atoms, the distances to the neighbouring atoms, and the disorder in the neighbour distances can be determined. The energy spectra for OEC in S_1 is shown in Figure 1.1.

1.3 Force Fields aka Potential Energy

Here will explain what force fields are and how they calculate a molecules potential energy.

1.4 Problem Definition

The goal of this thesis is to examine different search heuristics to determine the best method of finding the theoretical atomic structure of a molecule using the molecules EXAFS spectra for comparison. This problem contains two important but unrelated goals. Firstly, the algorithm must be able to find an atomic structure who's EXAFS spectra matches the experimental EXAFS spectra, and also create an atomic structure who's potential energy is a low as possible.

EXAFS can be used to identify properties of a molecule, but they do not provide enough detail to determine the atomic structure of a molecule in 3-dimensional space. An EXAFS spectra allows you to identify how far apart atoms are from each other, but does not give enough information to identify their dihedral angles. Fortunately, EXAFS can be used to assist in determining the atomic structure of a molecule. The energy spectra given off by the molecule is unique for its structure, meaning that you can create an atomic structure, obtain its EXAFS spectra, and compare the results. The hope is that if you create an atomic structure whose spectra closely matches the spectra of an actual model, then there is a high likelihood that the created structure will closely match the actual structure.

The IFEFFIT XAFS data analysis suite [3] is used to simulate the EXAFS experiments. FEFF6 is used to simulate an XAFS experiment and IFEFFIT does post processing of the simulated EXAFS spectra. During the atomic structure refinement, the generated atomic structures will be run through these applications to obtain an EXAFS spectra.

NAMD [4] will be used for the energy calculations. The NAMD Energy Plugin [5] will calculate the potential energy of the generated atomic structure.

Chapter 2

Previous Research

Before we can begin explaining the techniques we used in the next chapter it is necessary that we explain what research has already been done in this field.

2.1 Quantum Mechanics/Molecular Mechanics

In [6] Sproviero, Eduardo M and Gascón, José A and McEvoy, James P and Brudvig, Gary W and Batista, and Victor S used DFT-QM/MM and R-QM/MM techniques to find close approximations of the experimental EXAFS spectra. The EXAFS spectra used in their calculations was at a poorer resolution compared to the spectra used in the experiments in this work.

Density functional theory quantum mechanics/molecular mechanics (DFT-QM/MM) uses the atoms spatially dependent electron density (**CITE?**) to determine the position of each atom. Since DFT largely uses function approximations this approach is very limited.

To increase their accuracy the researchers used a refined quantum mechanics/molecular mechanics (R-QM/MM) technique. This approach iteratively adjusted the molecular structure of the molecule and attempted to minimize a scoring function defined in terms of the sum of squared deviations between the experimental and calculated EXAFS spectra. A quadratic penalty was applied to each atom to ensure that their

positions did not deviate too far from their original position in order to keep the energy of the system at a minimum.

The researches speculate that even though the R-QM/MM technique was able to generate an EXAFS spectra closer to the experimental spectra their solution was only a local solution because it was based on their original DFT-QM/MM solution.

In [7] Lubner, Sandra and Rivalta, Ivan and Umena, Yasufumi and Kawakami, Keisuke and Shen, Jian-Ren and Kamiya, Nobuo and Brudvig, Gary W and Batista, and Victor S repeated their original experiments performed in [6] with updated X-ray diffraction (XRD) data that had a resolution of 1.9Å. They have success rerunning the DFT-QM/MM experiment followed by the R-QM/MM experiment but still had the same speculations about remaining in a local solution.

Chapter 3

Background

The purpose of this chapter is to assist the reader in understand the search techniques used in this work.

3.1 Genetic Algorithm

A Genetic Algorithm (GA) is a search heuristic that is based on Darwin's theory of natural evolution. Darwin theorized that over a period of time a population of individuals would naturally mate and create offspring that were better than themselves. He suggests that not all individuals are created equally and that eventually the weaker individuals would die off. This same principle can be applied to a search algorithm. A GA contains a population of individuals that are evolved to find improved candidate solutions.

The individuals of a GA represent possible candidate solutions to the problem. Individuals are referred to as chromosomes. Typically a chromosome is represented as an array where each index of the array represents a property of the candidate solution. There are no restrictions to the encoding of a chromosome but each property of the chromosome must be independent from the others. The simplest example of a representation is a binary array of 1's and 0's, as shown in Figure 3.1.

Finding a representation is only the first step in creating a genetic algorithm. The

0	1	1	0	1	0	1	0
---	---	---	---	---	---	---	---

Figure 3.1: Simple Chromosome Representation

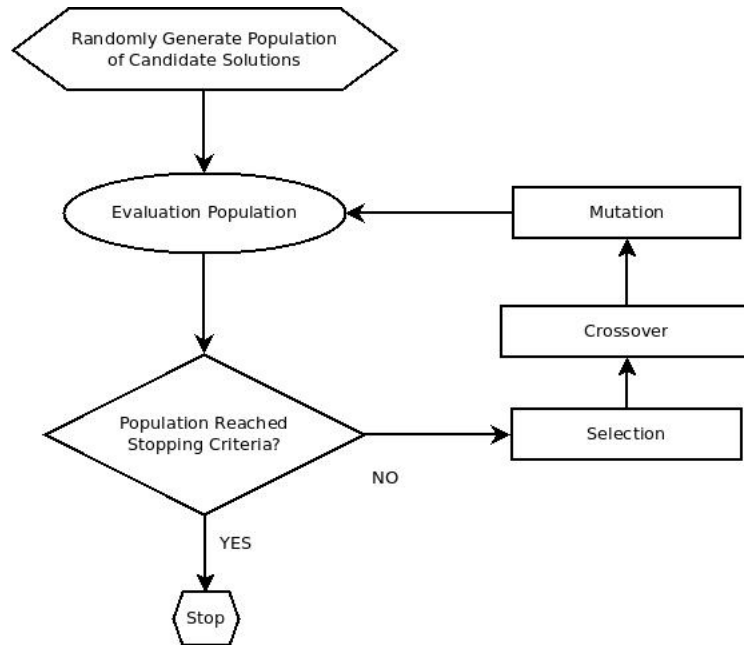


Figure 3.2: Population Individual Modification

next step is to define your set of operators. A GA consists of multiple operators that assist in the evolution of the population. Figure 3.2 depicts the flowchart of a genetic algorithm and where each genetic operation is performed.

3.1.1 Genetic Operators

:

Evaluation Function:

Selection Operator:

Crossover Operator:

Mutation Operator:

Elitism Operator:

3.2 Restering Genetic Algorithm

The RGA is a variation of the Recentering-Restarting Genetic Algorithm (RRGA) [8] [9] which has success in avoiding local minima. The RRGA is used to avoid fixating on local optima. RRGA does a series of GA runs which involves restarts and adjustments to the starting population. The RRGA selects a center, which is a possible candidate solution to the problem, and at the end of each basic GA run the center is compared to the best individual in the population. If the best individual is better than the current center it is replaced with the best individual and the whole process is repeated.

The RGA works similarly to the RRGA but there is no center for the population. Instead a basic GA is run until the population begins to converge. After a specified convergence percentage is reached, new individuals are added to the population. For example, if there are 100 individuals and the convergence rate is 5% then after all the duplicates are removed there will only be 5 individuals remaining. In restarting, shown in Algorithm 1, duplicate individuals were replaced with new individuals that have not yet entered the population. The implementation of restarting used in this work is described more fully in Section TBA.

Algorithm 1 Restarting the population

```

if population has converged to minimum diversity then
  remove all duplicate individuals;
  while population not full do
    insert random draw from generated individuals into population;
  end while
end if

```

3.3 Particle Swarm Optimization

3.4 Differential Evolution

Bibliography

- [1] D. J. Vinyard, G. M. Ananyev, and G. C. Dismukes, “Photosystem ii: The reaction center of oxygenic photosynthesis,” *Annual Review of Biochemistry*, vol. 82, pp. 577 – 606, 2013.
- [2] M. Newville, “Fundamentals of xafs,” *Consortium for Advanced Radiation Sources, University of Chicago (USA)*[<http://xafs.org>], 2004.
- [3] T. U. of Chicago, “Ifeffit: Interactive xafs analysis.” Accessed: 2014-04-01.
- [4] Theoretical and C. B. Group, “Namd - scalable molecular dynamics.” Accessed: 2014-04-01.
- [5] Theoretical and C. B. Group, “Namd energy plugin.” Accessed: 2014-04-01.
- [6] E. M. Sproviero, J. A. Gascón, J. P. McEvoy, G. W. Brudvig, and V. S. Batista, “A model of the oxygen-evolving center of photosystem ii predicted by structural refinement based on exafs simulations,” *Journal of the American Chemical Society*, vol. 130, no. 21, pp. 6728–6730, 2008.
- [7] S. Luber, I. Rivalta, Y. Umena, K. Kawakami, J.-R. Shen, N. Kamiya, G. W. Brudvig, and V. S. Batista, “S1-state model of the o2-evolving complex of photosystem ii,” *Biochemistry*, vol. 50, no. 29, pp. 6308–6311, 2011.
- [8] J. Hughes, S. Houghten, and D. Ashlock, “Recentering, reanchoring & restarting an evolutionary algorithm,” in *Nature and Biologically Inspired Computing (NaBIC), 2013 World Congress on*, pp. 76–83, IEEE, 2013.
- [9] J. Hughes, J. A. Brown, S. Houghten, and D. Ashlock, “Edit metric decoding: Representation strikes back,” in *Evolutionary Computation (CEC), 2013 IEEE Congress on*, pp. 229–236, IEEE, 2013.