Modeling Metal Protein Complexes from Experimental Extended X-ray Absorption Fine Structure using Evolutionary Algorithms

Collin Price Sheridan Houghten Sergey Vassiliev Doug Bruce

Brock University

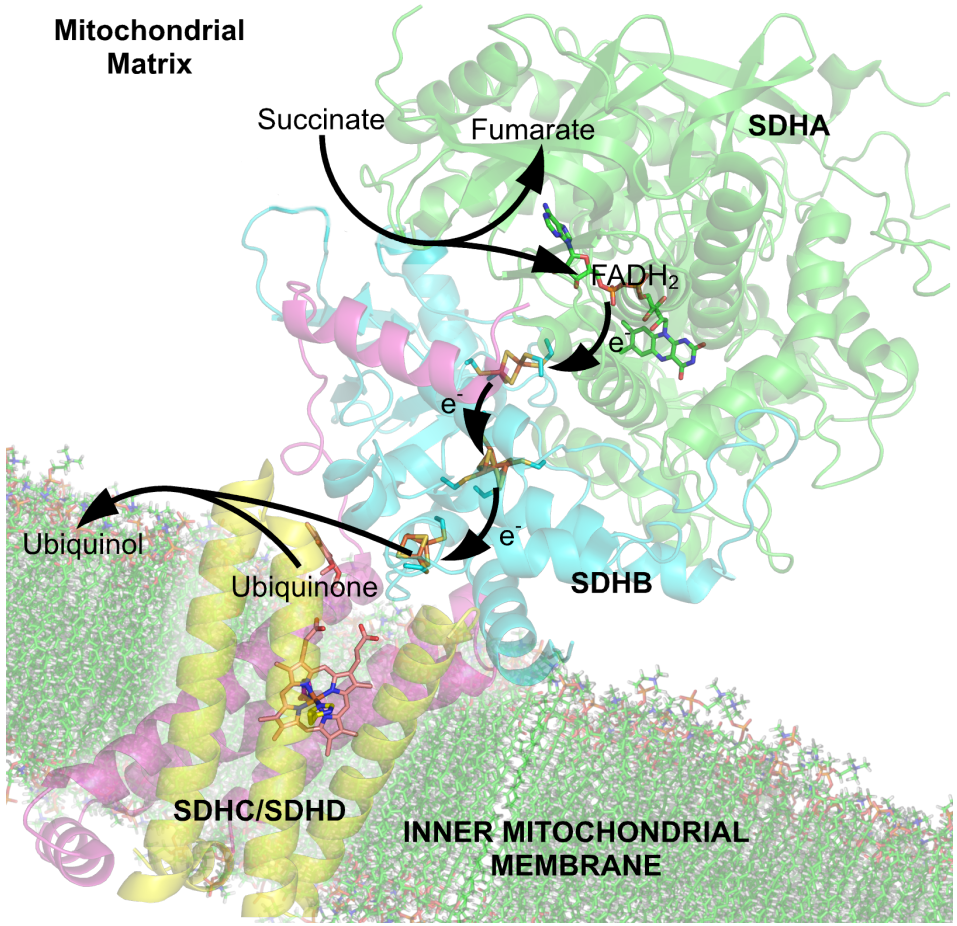
April 7, 2014

Outline

Cofactor

- "helper" molecules
- assist in biochemical transformations
- non-protein chemical compound
- exist within a protein

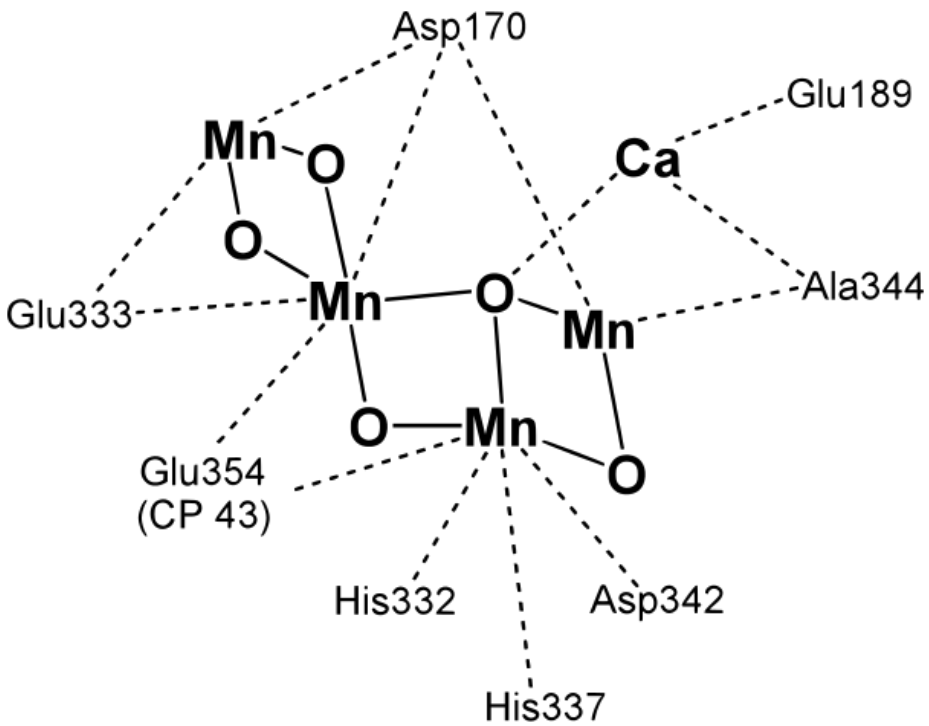
Figure : Cofactor Examples



Oxygen-Evolving Complex (OEC)

- Water-splitting complex.
- Exists in 5 states: S_0 to S_4.
- Splits H_2O into hydrogen and oxygen.
- S_4 reacts with water to produce oxygen.
- Process is not completely understood.
- S_1 is examined in this work.

Figure : Metalloenzyme core of the OEC



Experimental Extended X-ray Absorption Fine Structure (EXAFS)

- A method used to measure the absorption coefficient of a material as a function of energy.
- X-ray is tuned to have the same wavelength as the target atom.
  - In the case of the OEC we targeted the four Manganese atoms.
- X-ray reacts with target atom causing the atom to lose an electron and interact with neighbouring atoms.
- Fluorescence energies are emitted and measured.

Experimental EXAFS Spectra

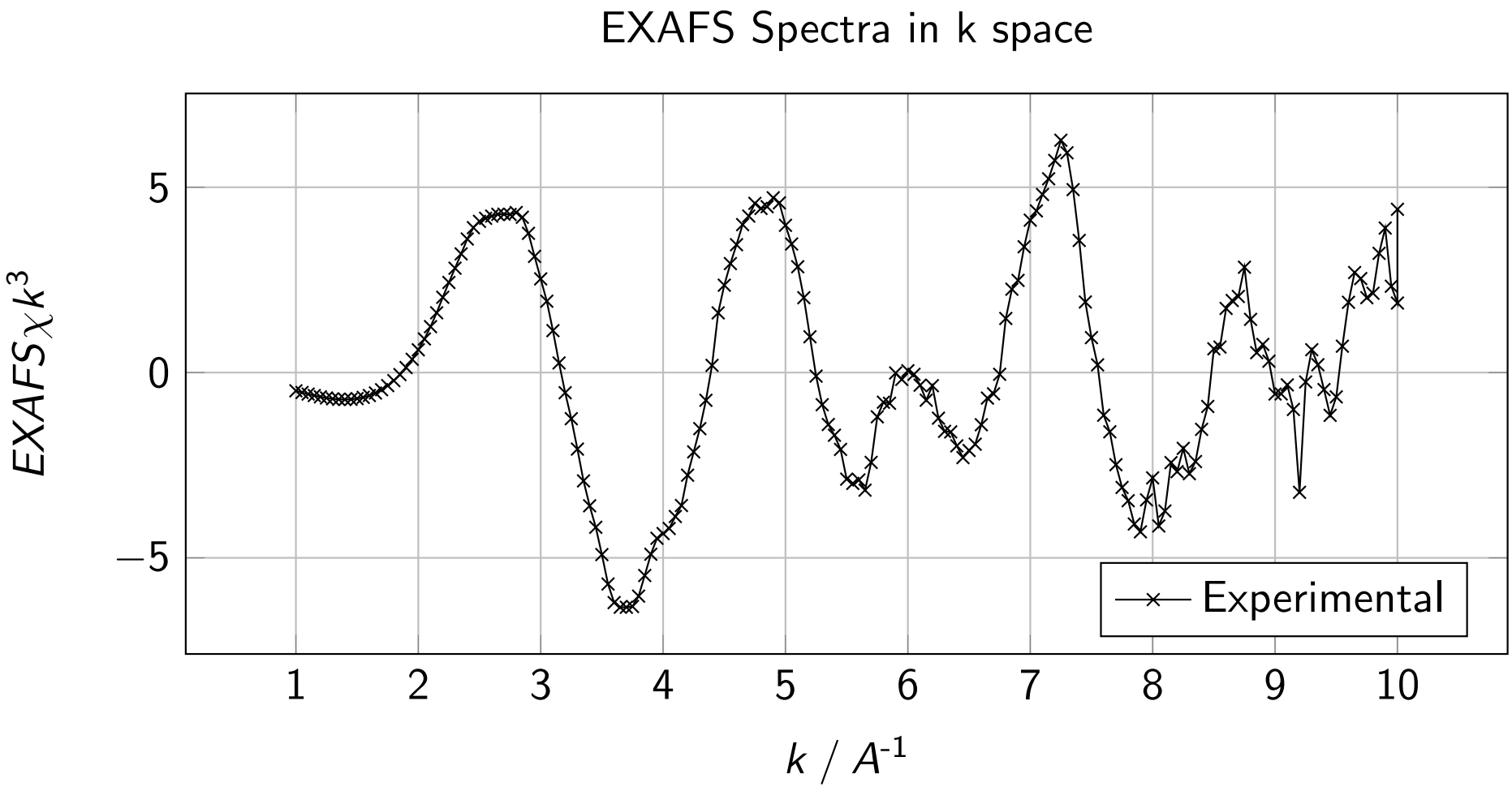


Figure : EXAFS Spectra of OEC in S_1

Structure Refinement Problem

- EXAFS spectra provides limited information about neighbouring atoms.
- We would like to know the 3-dimensional positions of each atom relative to each other.
- EXAFS spectra can be generated in simulations.
- If we match the calculated EXAFS spectra with the experimental EXAFS spectra it is likely we found the correct configuration.
- Goal is to minize the differences between spectra.

Example Spectra Comparison

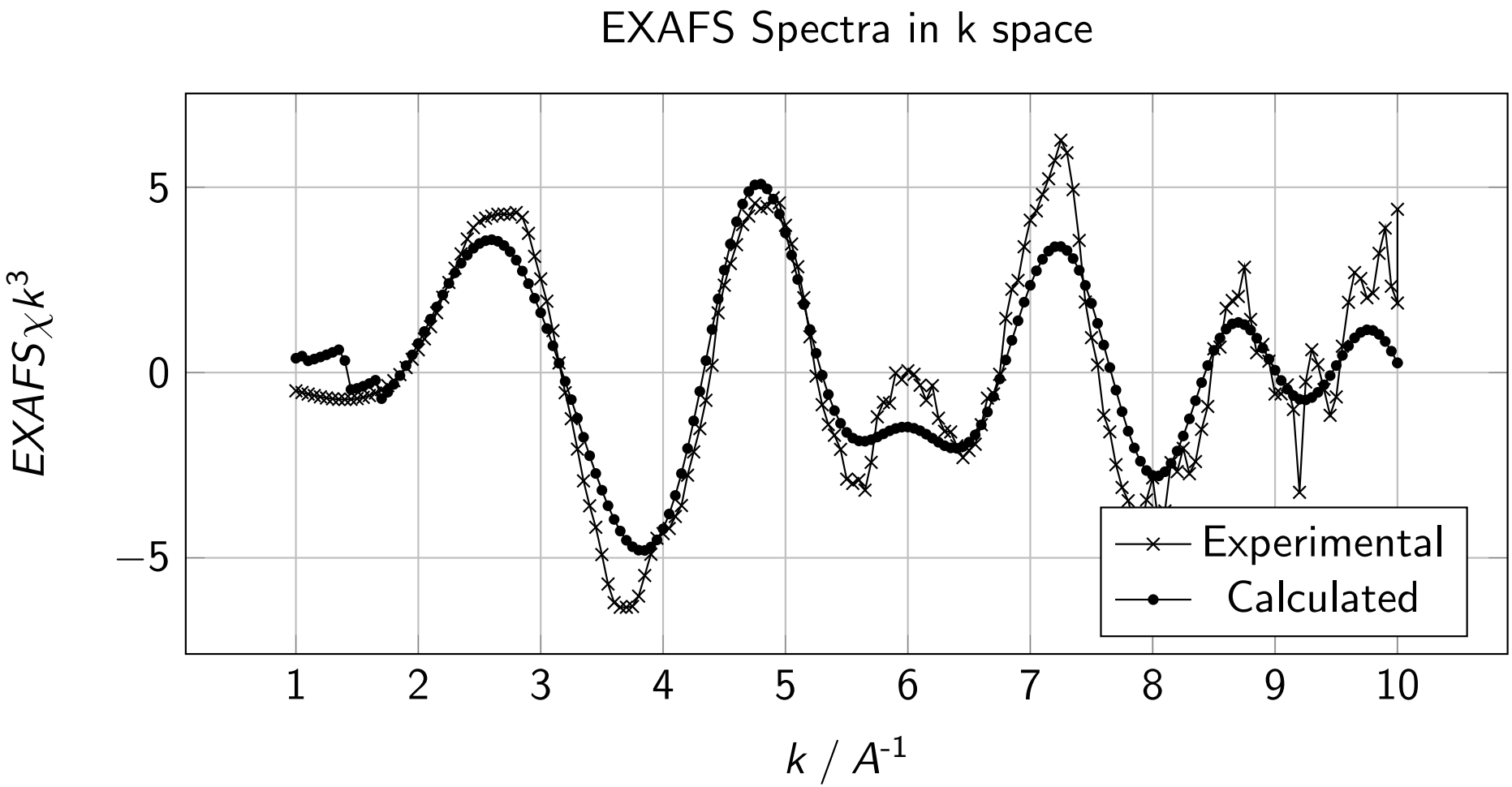


Figure : OEC EXAFS Spectra Comparison

Previous Work

- A model of the oxygen-evolving center of photosystem II predicted by structural refinement based on EXAFS simulations
  - Sproviero, Eduardo M and Gascón, José A and McEvoy, James P and Brudvig, Gary W and Batista, Victor S
  - Quantum Mechanical and Molecular Mechanical hybrid
- S_1-state model of the O_2-evolving complex of photosystem II
  - Luber, Sandra and Rivalta, Ivan and Umena, Yasufumi and Kawakami, Keisuke and Shen, Jian-Ren and Kamiya, Nobuo and Brudvig, Gary W and Batista, Victor S
  - Used a conjugate gradient optimization technique
  - Monte Carlo
- Previous techniques would converge on local minima.

Genetic Algorithm (GA)

- Stochastic metaheuristic that mimics the process of natural evolution.
- A population of candidate solutions to an optimization problem are evolved toward better solutions.
- New candidates are created using crossover and mutation operators on existing candidate population.

Restarting Genetic Algorithm (RGA)

- Variation of the Recentering-Restarting Genetic Algorithm.
- Used to maintain diversity within the population.
- New candidate solutions are injected into the population once the convergence criteria is met.

Restarting Genetic Algorithm

Algorithm 1 Restarting the population

```
1: if population has converged to minimum diversity then
2:     remove all duplicate individuals;
3:     while population not full do
4:         insert random draw from generated individuals into population;
5:     end while
6: end if
```

Representation

- Set of 3-dimensional coordinates in space, where each position is assigned to a specific atom in the atomic structure.
- Entire OEC atomic structure contains 1269 atoms.
- EXAFS calculations only require 79 specific atoms.
- The units of measurement for each atom position are measured in Angstroms (Å).

Table : Example Chromosome

x_0	y_0	z_0
x_1	y_1	z_1
x_2	y_2	z_2
x_3	y_3	z_3
...	...	...

Population

- Initial OEC atomic structure came from the crystallographic photosystem II structure in the Protein Data Bank.
- Randomly moving each atom produced too many unusable configurations.
- Needed to generate feasible structures for starting GA population.
- Used a molecular dynamics simulation to obtain these structures.
- Produces 10 000 atomic structures.
- Evaluated each and kept the top 300.

Fitness

- The goal of the experiment is to find an atomic structure that generates the same EXAFS spectra as the experimental EXAFS spectra.
- FEFF6 is used to simulate an XAFS experiment.
- IFEFFIT does post processing of the simulated EXAFS spectra.
- The root-mean-square deviations (RMSD) is used to calculate the difference between the experimental and calculated EXAFS spectra.

$$RMSD = \sqrt{\frac{\sum_{t=1}^{n} \left(x_{1,t} - x_{2,t}\right)^2}{n}} \tag{1}$$

Example Spectra Comparison

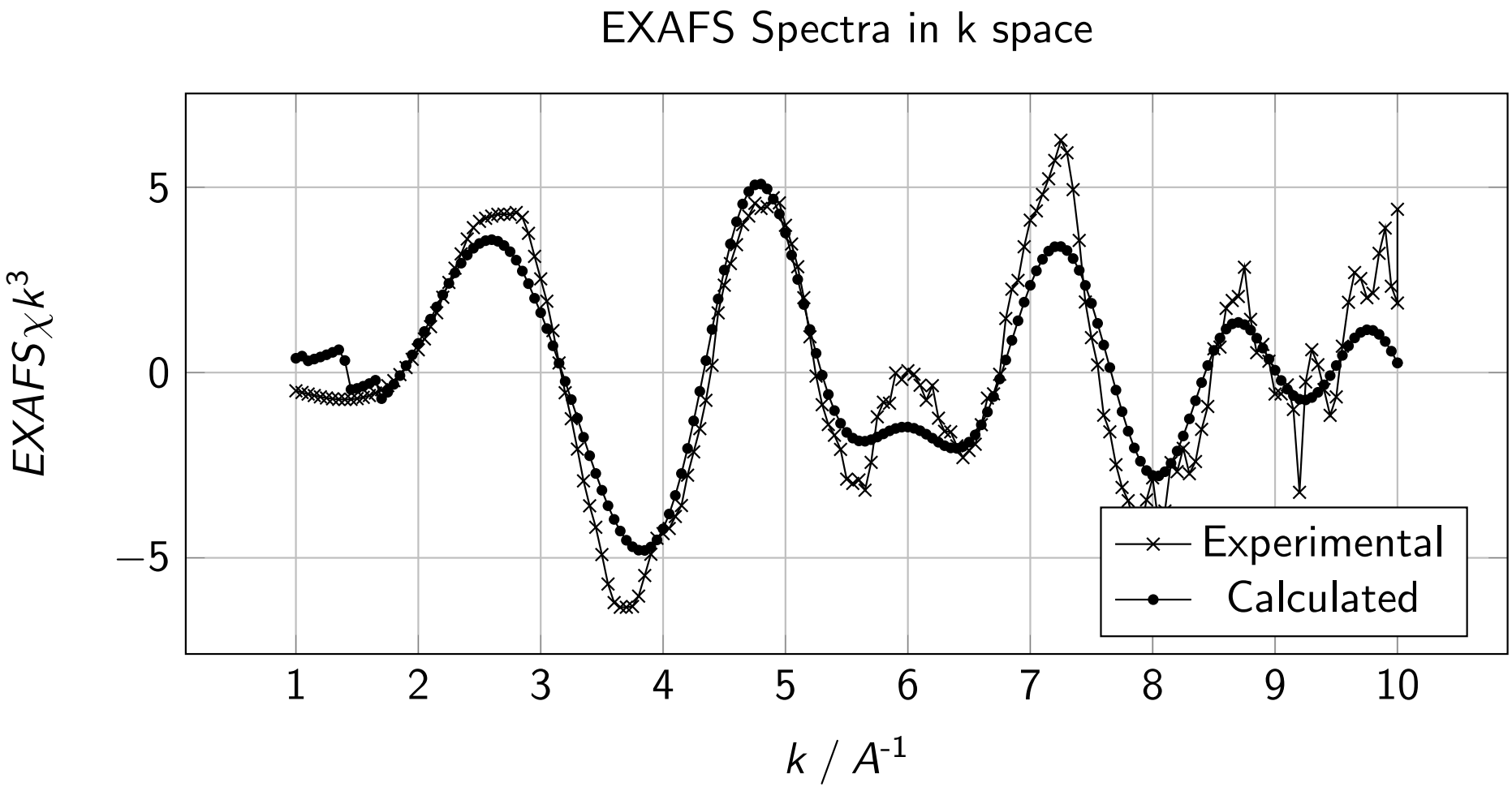


Figure : OEC EXAFS Spectra Comparison

Operators

- Crossover operator
  - One-point crossover
- Mutation operator
  - Single atomic coordinate is randomly moved my 0.05Å.
  - Tested moving atomic coordinates by 0.001Å, 0.005Å, 0.01Å, 0.025Å, 0.05Å, 0.1Å, 0.5Å, 1Å, and 5Å.
  - 0.05Å was chosen because it produced at least a 1% change in fitness for all chemical elements.
- Selection operator
  - 3-tournament selection

Experiment Setup

Table : System Parameters used in GA and RGA

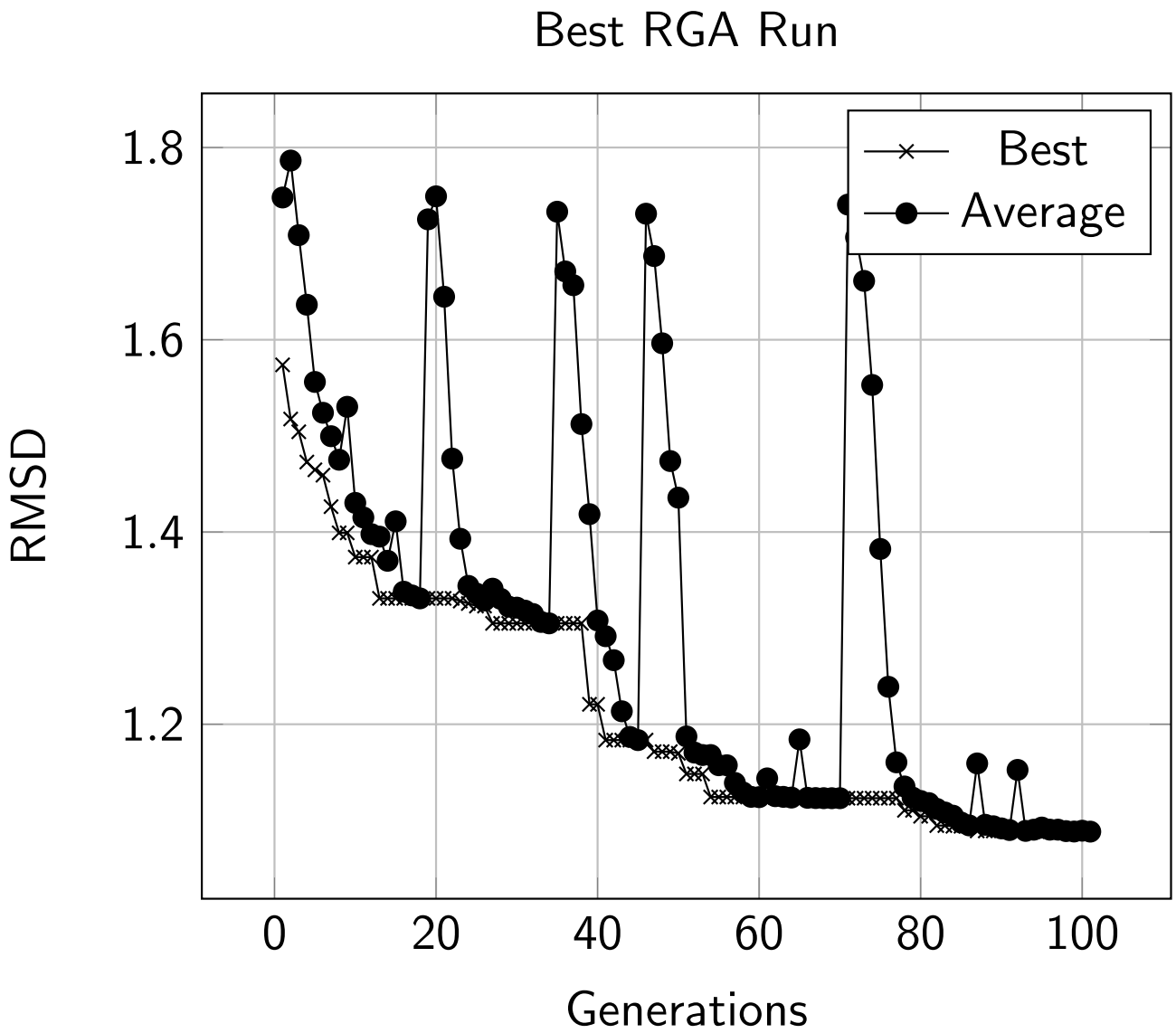
Settings	Values
Runs	10
Generations	Max. 30, Until Converged
Population Size	50
Crossover Rate	80, 70
Mutation rate	10, 20, 30
Elitism	True, False
Number of restart attempts	3, 5
Max convergence percentage before restarting	5%, 10%

Results from GA and RGA Experiments

Gen.	Crossover	Mutation	Elitism	Conv. Rate	Restarting	Avg. Best
30	80%	20%	False	-	-	1.3411
30	80%	10%	False	-	-	1.3746
30	70%	30%	False	-	-	1.2756
30	80%	10%	True	-	-	1.3278
30	70%	30%	True	-	-	**1.2591**
30	80%	20%	True	-	-	1.3299

Gen.	Crossover	Mutation	Elitism	Conv. Rate	Restarting	Avg. Best
67	80%	20%	True	10%	3	1.2247
81	80%	20%	True	5%	3	1.2117
107	80%	20%	True	10%	5	1.2062
124	80%	20%	True	5%	5	1.2156
67	70%	30%	True	5%	5	**1.1998**
81	70%	30%	True	10%	5	1.2138

Results


Best RGA Run
Best
Average
RMSD
1.8
1.6
1.4
1.2
0
20
40
60
80
100
Generations

Best RGA Spectra

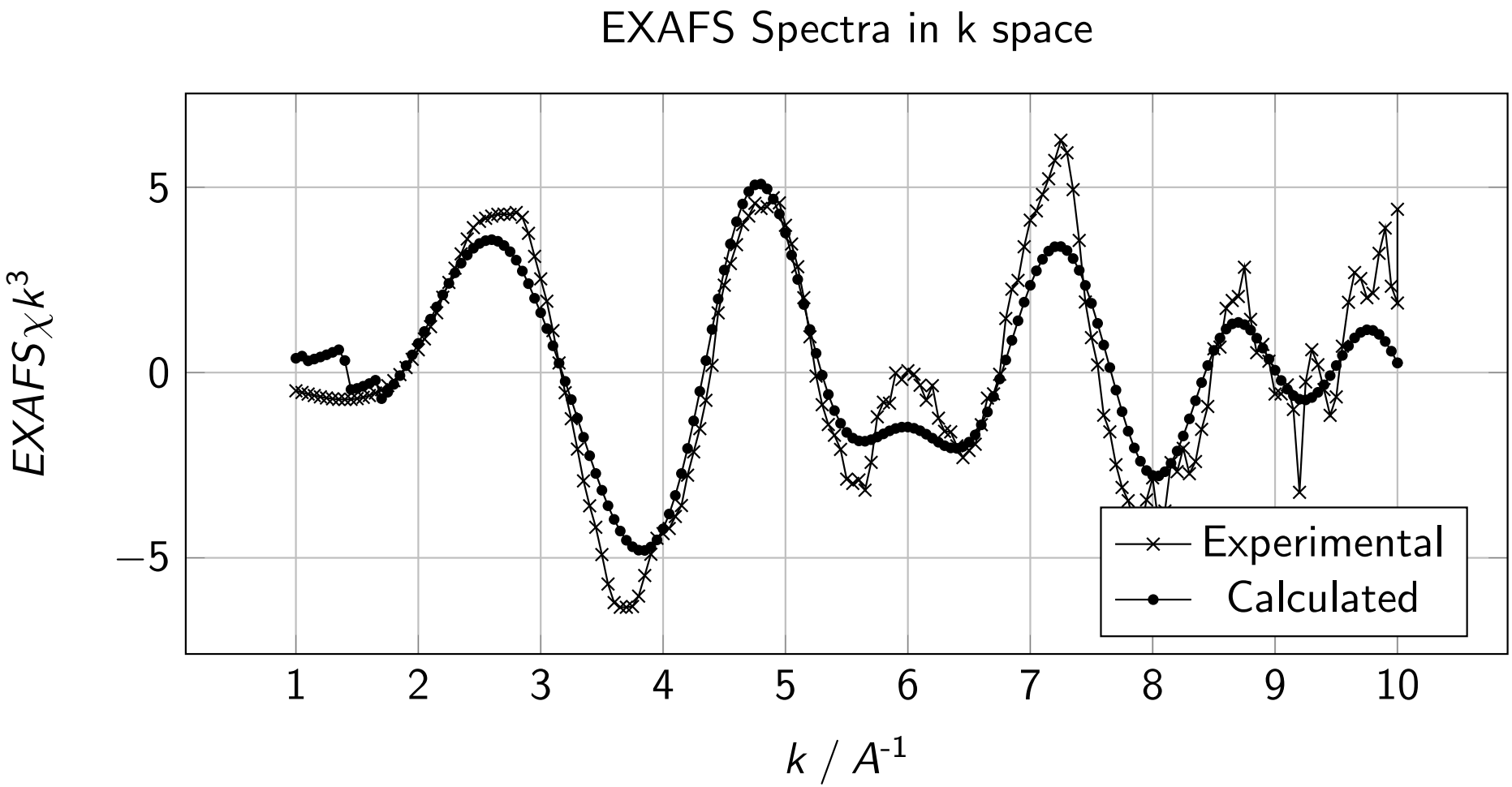


Figure : OEC EXAFS Spectra Comparison

Results

- GA and RGA were able to find a better solution than previously found.
- RGA was able to find a better solution than GA.

Table : Comparison of Results

Algorithm	Best Results
DFT-QM/MM	1.2643
R-QM/MM	1.2403
GA	1.1452
RGA	**1.0877**

Post-Optimization

- Results from GA and RGA were successful but we could do better.
- Convinced by guest speaker to try Particle Swarm Optimization.
- Convinced by fellow student to try Differential Evolution.
- Used the computational intelligence library, CILIB.

Particle Swarm Optimization (PSO)

- Better suited for continuous space problems.
- Contains a population of candidate solutions called particles.
- Particles move around the search space.
- Particles positions are updated based on the global best position and their local best position.

Differential Evolution (DE)

- Better suited for continuous space problems.
- Contains a population of candidate solutions called agents.
- An agents position is combined with other agents positions.
- If new agent is better, the old one is discarded.

Population

- Populations were initialized the same for both PSO and DE.
- Starting atomic structure was best found from RGA experiments.
- Generated population by randomly moving atomic coordinates by ± 0.05Å and ± 0.25Å.

Results

Table : Local Search Results

Algorithm	Initial Movement Radius	Pop. Size	Gen.	Average Best
PSO	±0.05Å	50	30	0.8976
DE	±0.05Å	50	30	**1.1354**
PSO	±0.25Å	50	30	1.2411
DE	±0.25Å	50	30	1.7220
PSO	±0.05Å	50	100	**0.7782**

- DE was not able to improve upon previously found candidate solutions.

Best Result

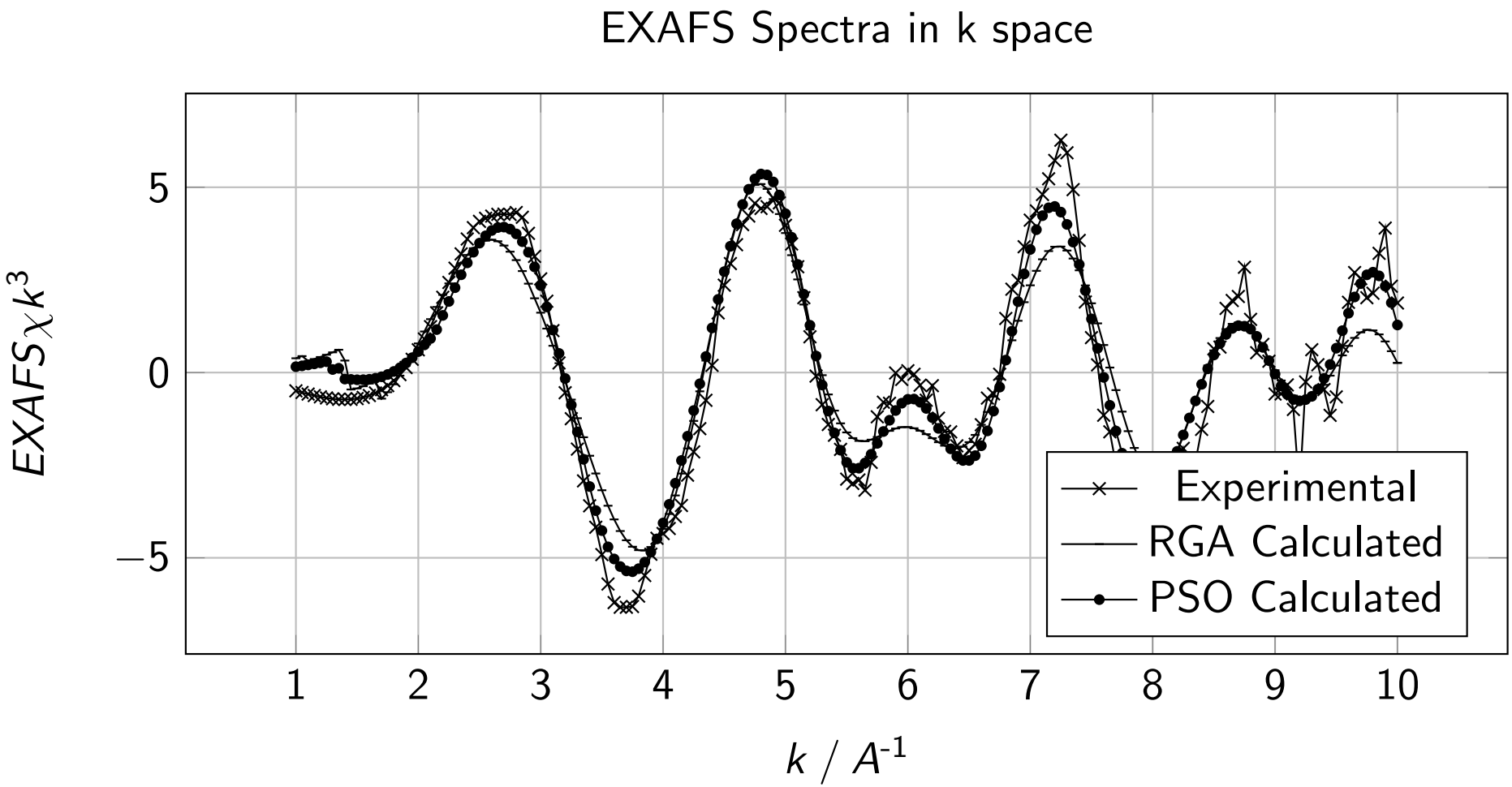


Figure : OEC EXAFS Spectra Comparison for PSO Post-Optimization

Final Results

Table : Comparison of Results

Algorithm	Best Results
DFT-QM/MM	1.2643
R-QM/MM	1.2403
RGA	1.0877
PSO	**0.7287**

Subset Testing

- Individuals contain 79 atoms.
- Wanted reduce the search space.
- Tested keeping certain chemical elements rigid during experiments.

Table : Chemical Element Breakdown of OEC

Element	Count
Mn	4
Ca	1
O	26
C	14
N	6
H	28
Total	79

GA Parameters

Table : GA Subset Parameters

Runs	10
Population size	50
Crossover rate	0.7
Mutation rate	0.3
Elitism	True
Number of restart attempts	5
Max convergence percentage before restarting	5%

Results

Table : Results from subsets experiments

Atoms	Best	Average
Mn, Ca, C, O, N, H	1.1998	1.2580
Mn, Ca, C, O, N	**1.1697**	1.2621
Mn, Ca, C	2.4413	N/A
Mn, Ca, O	1.2531	N/A
Mn, Ca, N	2.4881	N/A
Mn, Ca, C, O	**1.1648**	N/A
Mn, Ca, C, N	2.5143	2.5649
Mn, Ca, O, N	N/A	N/A
Mn, Ca	2.4916	2.5088

Note: N/A means there were too many invalid solutions.

Future Work

- Attempt direct comparison of PSO and DE with GA and RGA.
- Perform another molecular dynamics simulation on best result found from experimentation.
- Incorporate force fields into the algorithms to guarantee chemically feasible atomic structures.
- Multi-objective optimization using RMSD and potential energy.

Questions?