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

Collin Price
Supervisor: Sheridan Houghten

Brock University
Computer Science

October 6, 2014

Outline

Cofactor

- "Helper" molecules
- Assist in biochemical transformations
- Non-protein chemical compound
- Exist within a protein

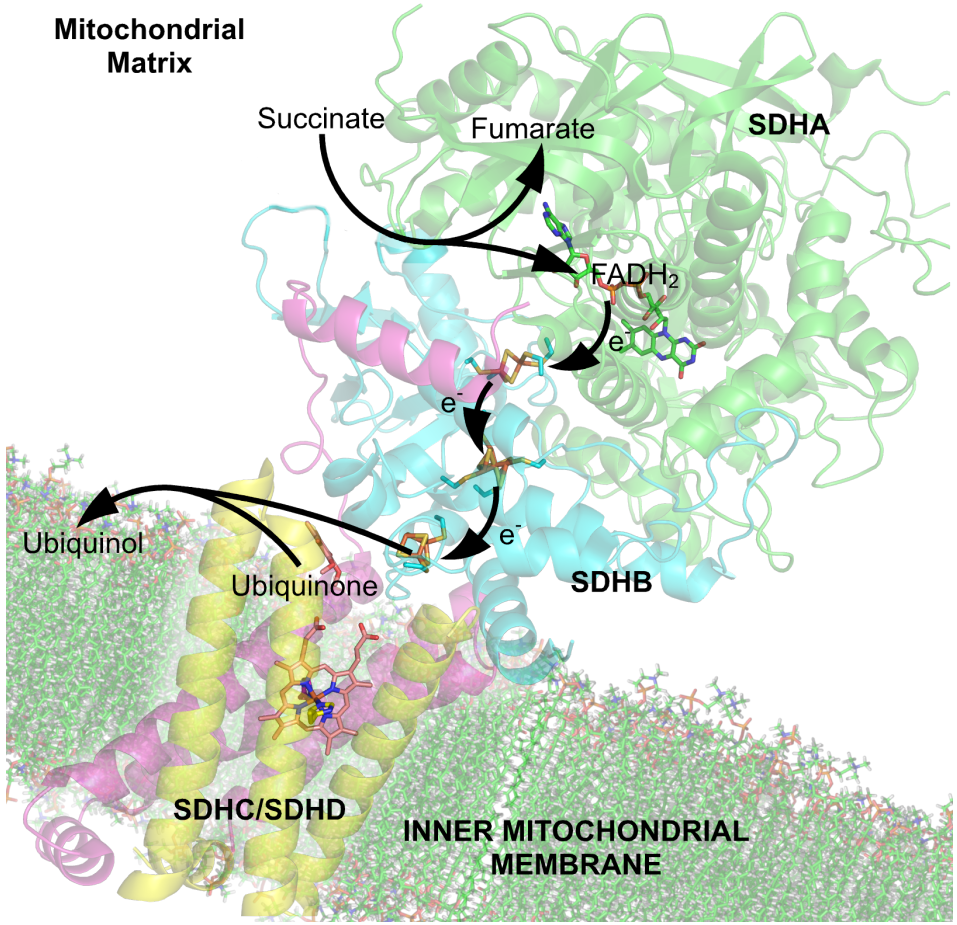


Figure : Cofactor Examples

Oxygen-Evolving Complex (OEC)

- The water-oxidizing enzyme of photosystem II.
- 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.

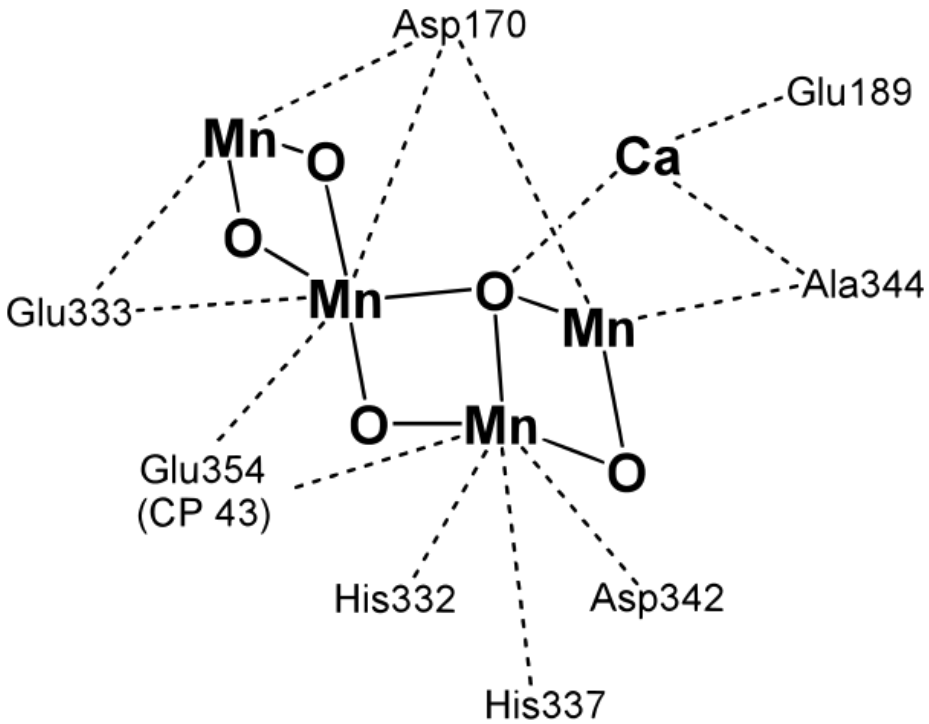


Figure : Metalloenzyme core of the OEC

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

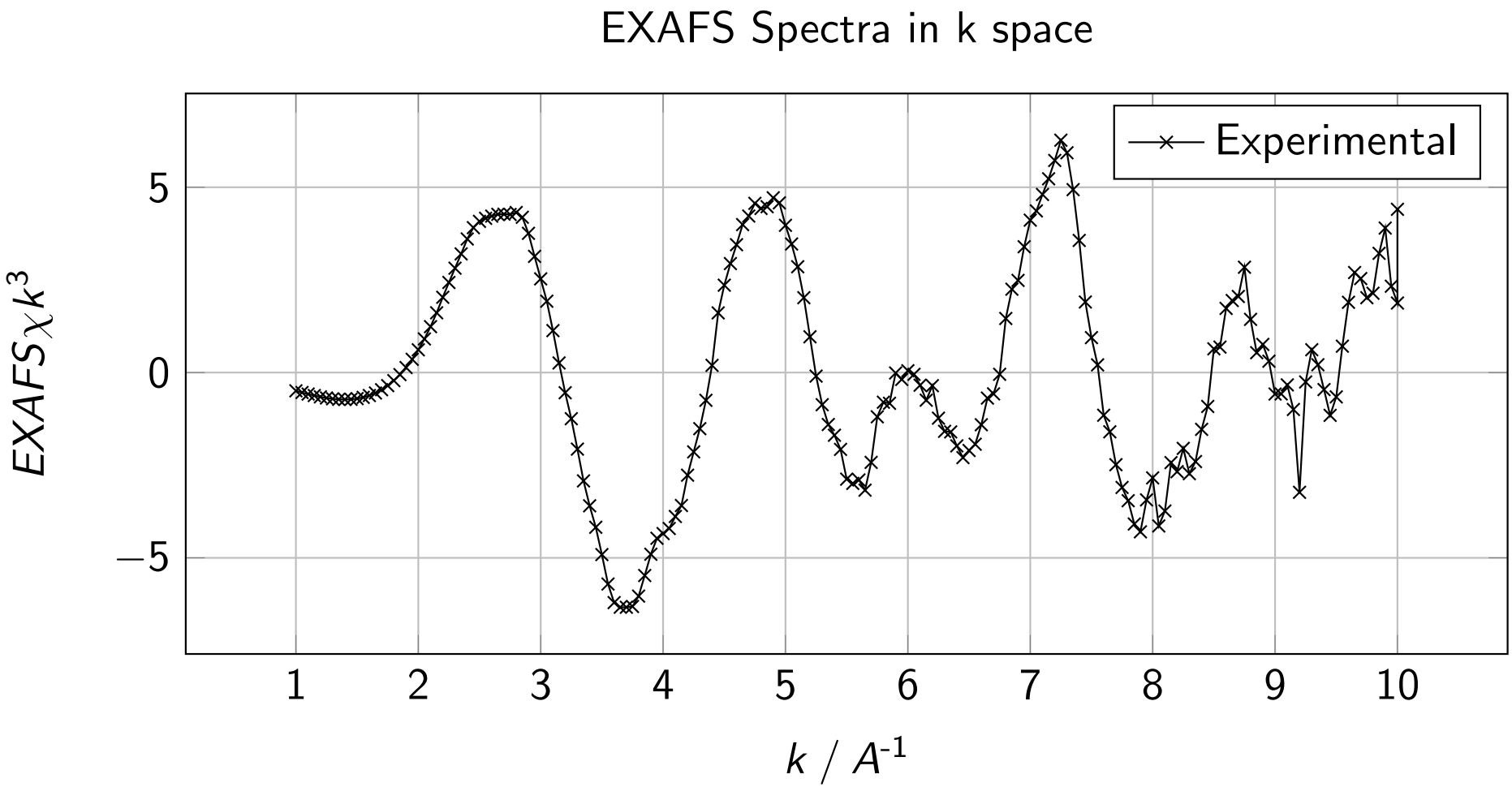


Figure : EXAFS Spectra of OEC in S_1

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 Spectrum Comparison

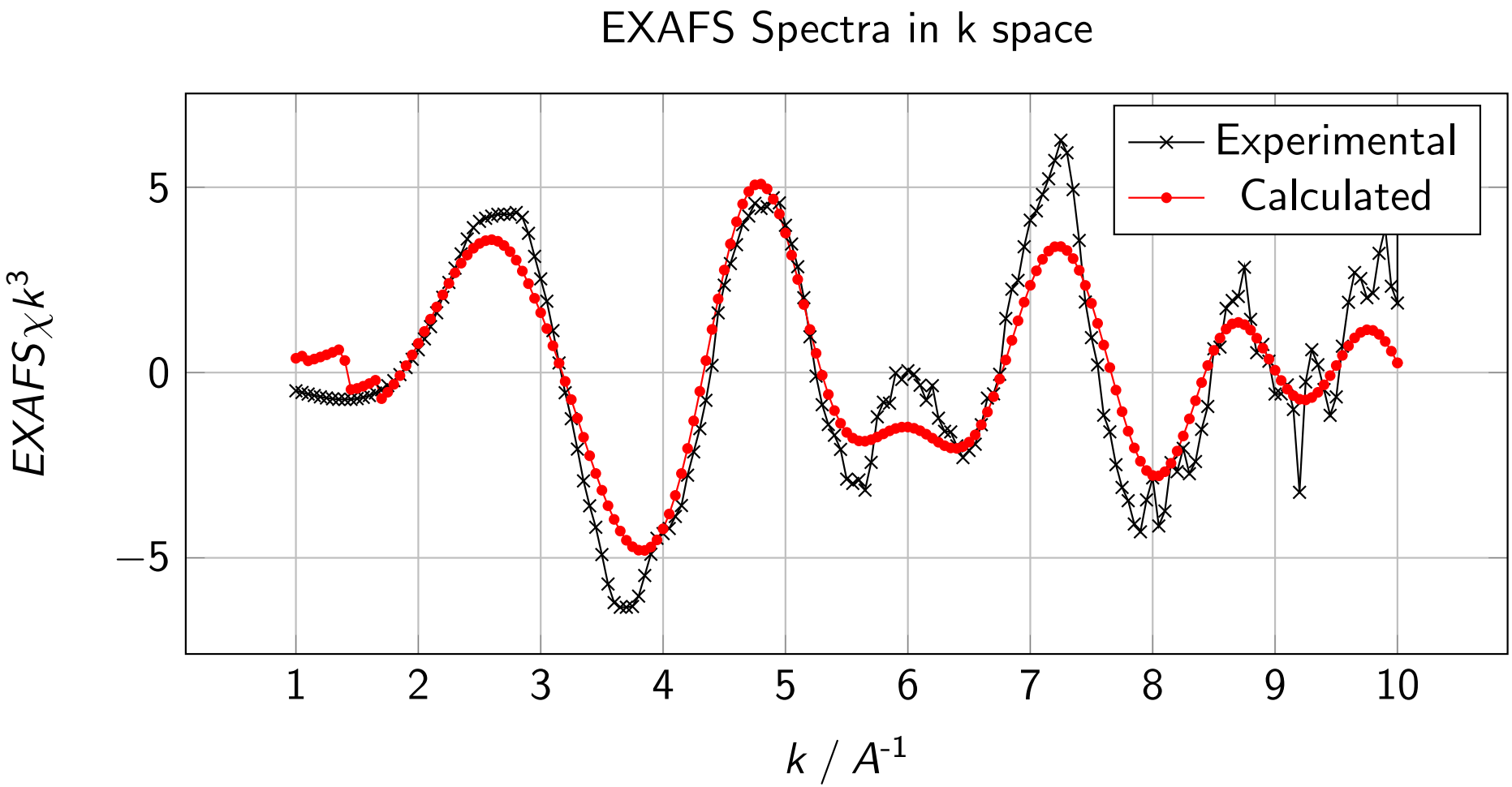


Figure : OEC EXAFS Spectrum Comparison

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*
- 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*
- Density functional theory quantum mechanics and molecular mechanics hybrid (DFT-QM/MM)
- Monte Carlo method (R-QM/MM)
- Previous techniques would converge on local minima.

Genetic Algorithm (GA)

- Shown success in past studies in finding low-energy protein conformations.
- Higher potential of avoiding local optima.
- Algorithm is able to produce a population of candidate solutions for future examination.

Restarting Genetic Algorithm (RGA)

- Variation of the Recentering-Restarting Genetic Algorithm.
- Used to maintain diversity within the population.
- Genetic Algorithm is run until the population has reached a set convergence rate.
- Duplicate individuals are removed from the population and new individuals are injected into the population.

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
```

Problem Encoding

- 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).

X	Y	Z
14.451	-13.346	1.133
15.336	-13.488	2.014
13.005	-13.364	1.452
0.019	0.011	0.045
...	...	...

Figure : Representation 1

Population Generation

- 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 (NAMD).
- Produced 10 000 atomic structures.
- Evaluated each atomic structure and kept the top 3%.

Operators

- Crossover operator
  - One-point crossover
  - Least destructive to the individuals
- Mutation operator
  - Single atomic coordinate is randomly moved my 0.05Å using Euclidean distance.
  - 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

Fitness

- The goal of the experiment is to find an atomic structure that generates the same EXAFS spectrum as the experimental EXAFS spectrum.
- FEFF6 is used to simulate an XAFS experiment.
- IFEFFIT does post processing of the simulated EXAFS spectra.
- IFEFFIT was developed at CARS, the Consortium for Advanced Radiation Sources at The University of Chicago.
- 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 Spectrum Comparison

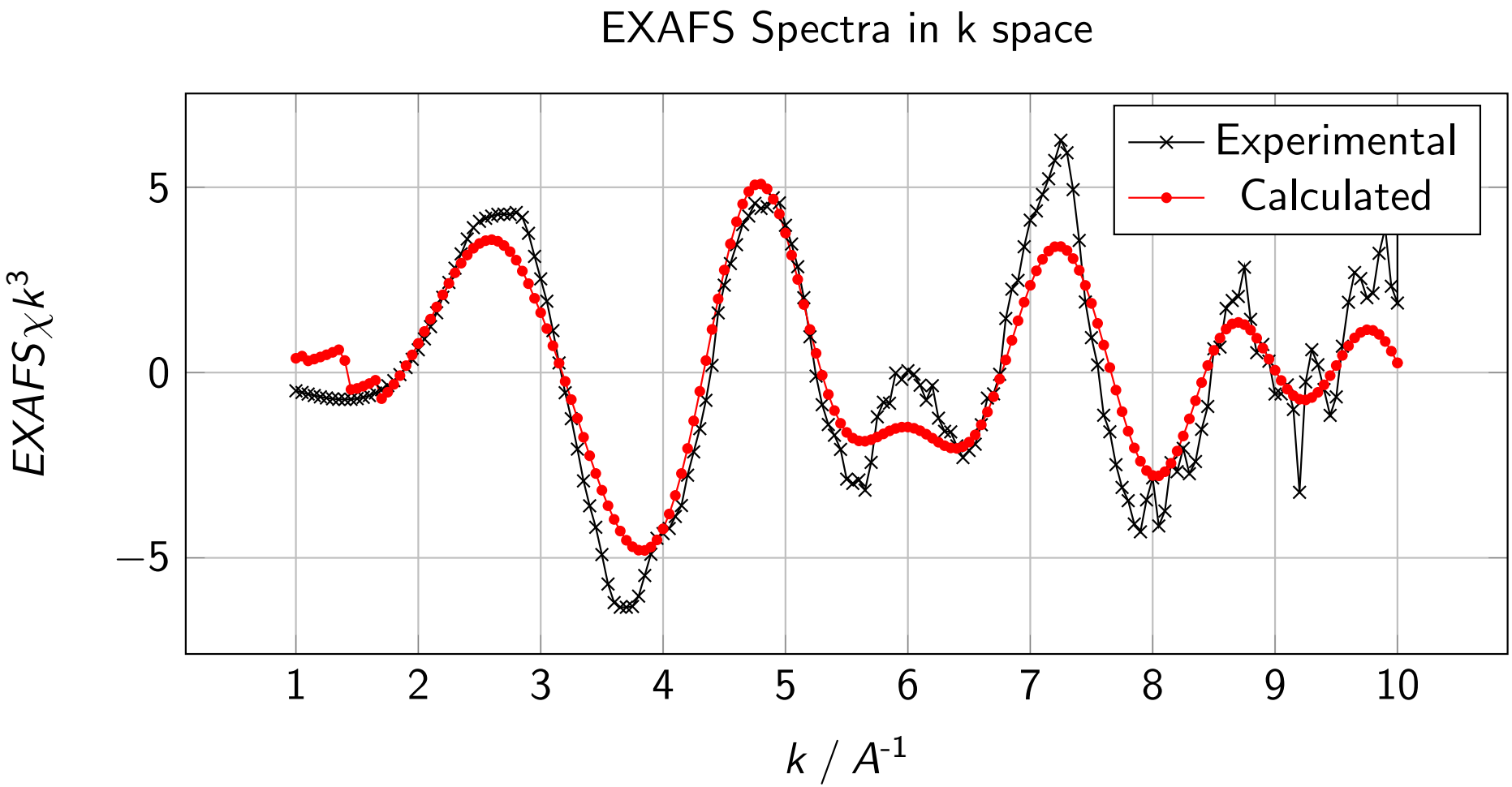


Figure : OEC EXAFS Spectrum Comparison

Experiment Setup

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

Table : System Parameters used in GA and RGA

Gen.	Crossover	Mutation	Elitism	Best RMSD	Avg. Best RMSD
30	80%	20%	False	1.2471	1.3518
30	80%	10%	False	1.1880	1.3610
30	70%	30%	False	1.1173	**1.2942**
30	80%	20%	True	1.2287	1.3294
30	80%	10%	True	1.2349	1.3658
30	70%	30%	True	**1.0533**	1.3044

Table : Results from GA Experiments

Gen.	Xover	Mut.	Elitism	Conv. Rate	Restarting	Best	Avg. Best
61	80%	20%	True	10%	3	1.1297	1.2532
73	80%	20%	True	5%	3	1.1174	1.2468
86	80%	20%	True	10%	5	1.0388	1.2252
106	80%	20%	True	5%	5	**0.9649**	1.2149
72	70%	30%	True	10%	3	1.1170	1.2229
83	70%	30%	True	5%	3	1.0012	1.2119
100	70%	30%	True	10%	5	1.0353	**1.1808**
133	70%	30%	True	5%	5	0.9992	1.1856

Table : Results from RGA Experiments

Analysis

Genetic Algorithm

- Mann-Whitney U tests performed.
- Experiments using a crossover rate of 70% and a mutation rate of 30% performed statistically better than others.
- May indicate problem prefers exploration over exploitation.
- Results show GA experiments converging early on local optima.

Restarting Genetic Algorithm

- Mann-Whitney U tests performed.
- RGA experiments performed statistically better than all GA experiments.
- Experiment with the highest number of restarts and lowest convergence criteria performed statistically better than others.

Best RGA Spectrum

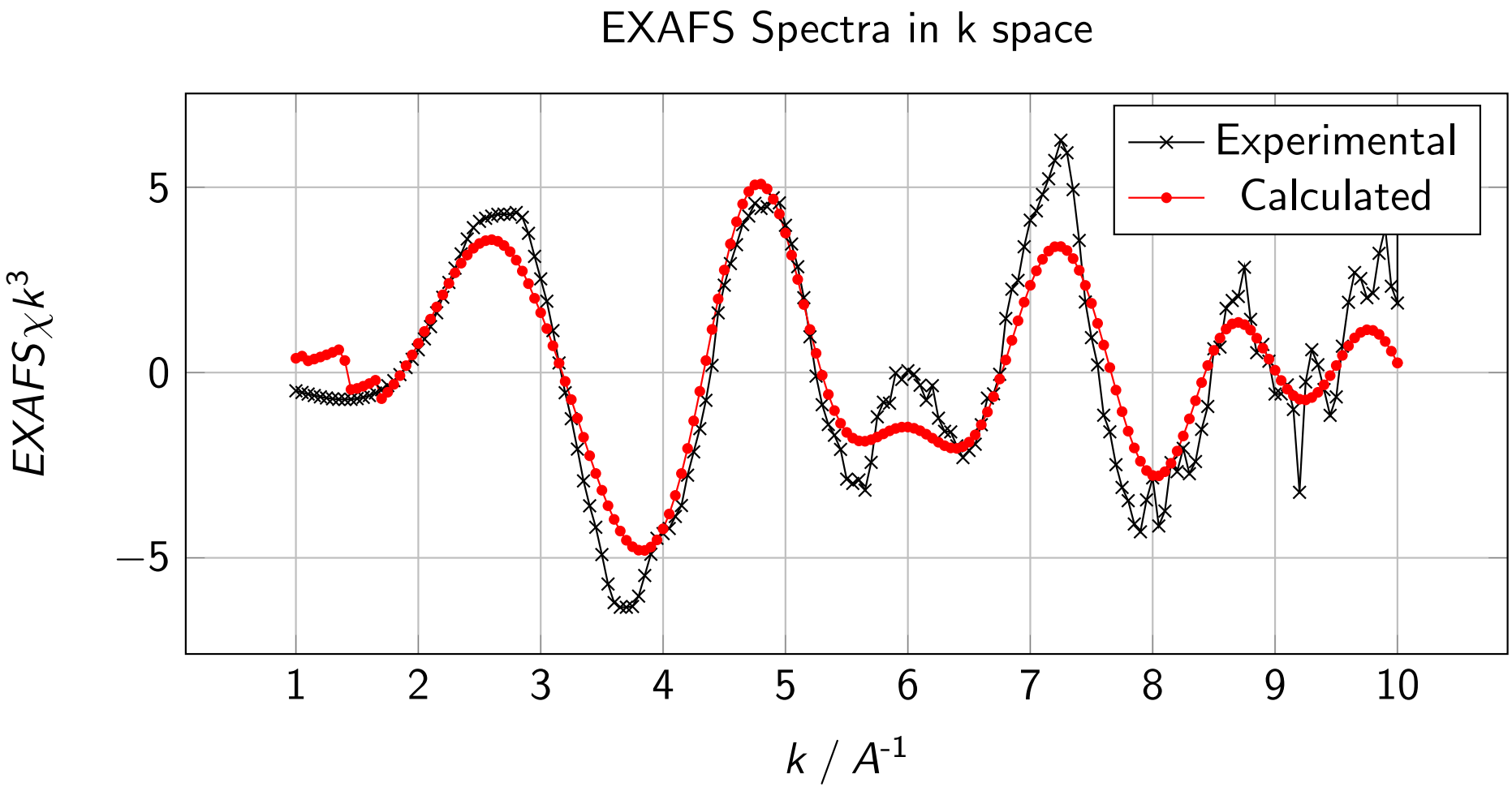
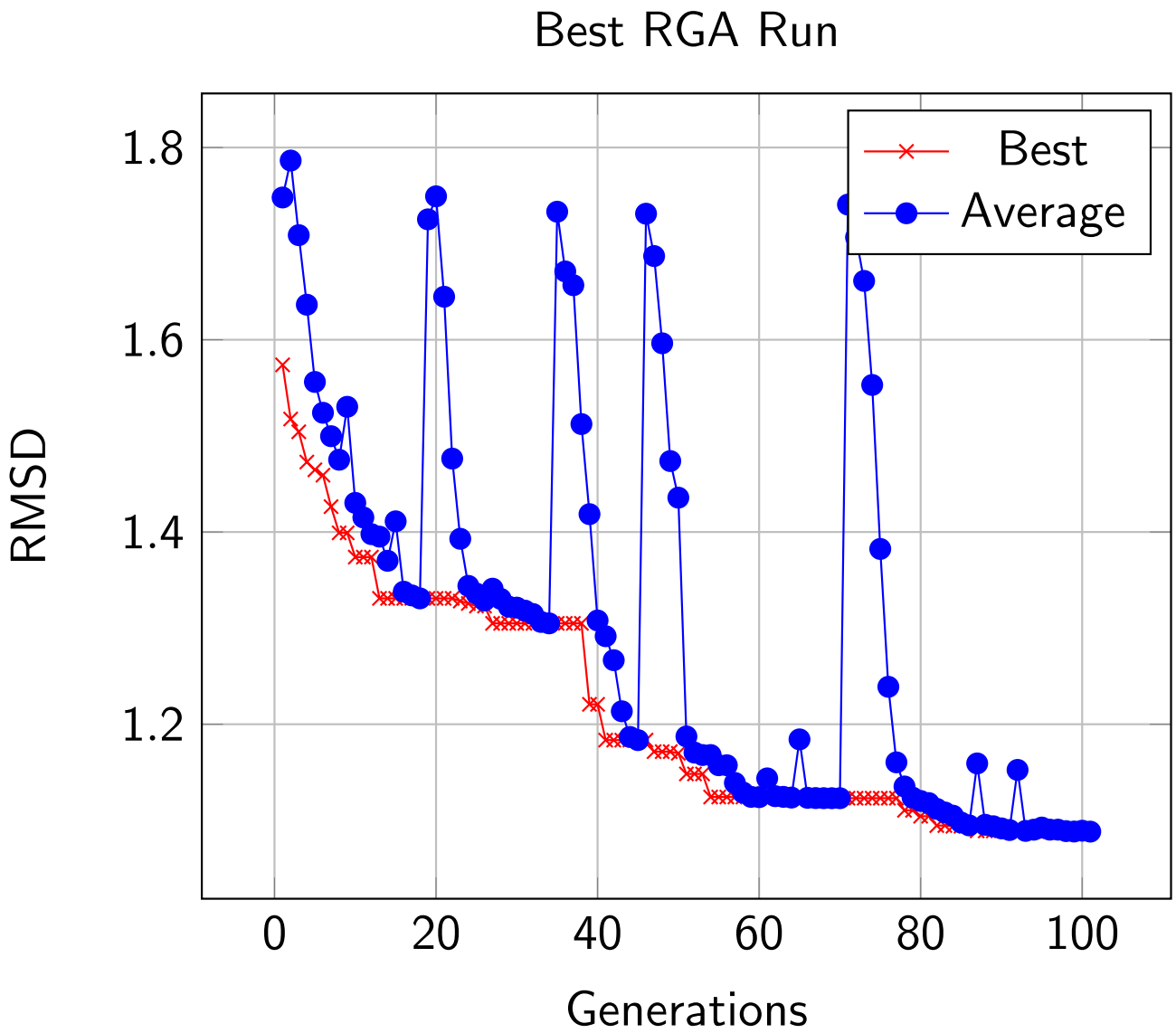


Figure : OEC EXAFS Spectrum Comparison

Results


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

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) & Differential Evolution (DE)

- Excel in continuous space problems.
- Offers more fluid movement through the search space.
- Still gain a population of candidate solutions.

Population Generation

- Molecular dynamics simulation used for GA/RGA could not be used for PSO and DE.
- Initial atomic structure was taken from the best candidate solution from GA/RGA experiments.
- Population was generated by randomly moving atomic coordinate of each atom by ± 0.05Å and ± 0.25Å.
- Populations were initialized the same for both PSO and DE.

Experiment Setup

Settings	Values
Runs	30
Generations	30
Population Size	50
Initial Movement Radius	0.05, 0.25

Table : System Parameters used in DE and PSO

Results

Algorithm	Initial Movement Radius	Best RMSD	Average Best RMSD
DE	±0.25Å	1.4118	1.7267
PSO	±0.25Å	0.9296	1.2445
DE	±0.05Å	0.9973	1.1386
PSO	±0.05Å	**0.7977**	**0.9001**

Table : Results for Post-Optimization DE, and PSO

- Initial RMSD from RGA experiment was 1.0877.
- DE was not able to improve upon previously found candidate solutions.
- Lower initial movement radius performed statistically better.
- Post-Optimization using PSO performed statistically better RGA.

Best Result: Post-Optimization PSO vs. RGA

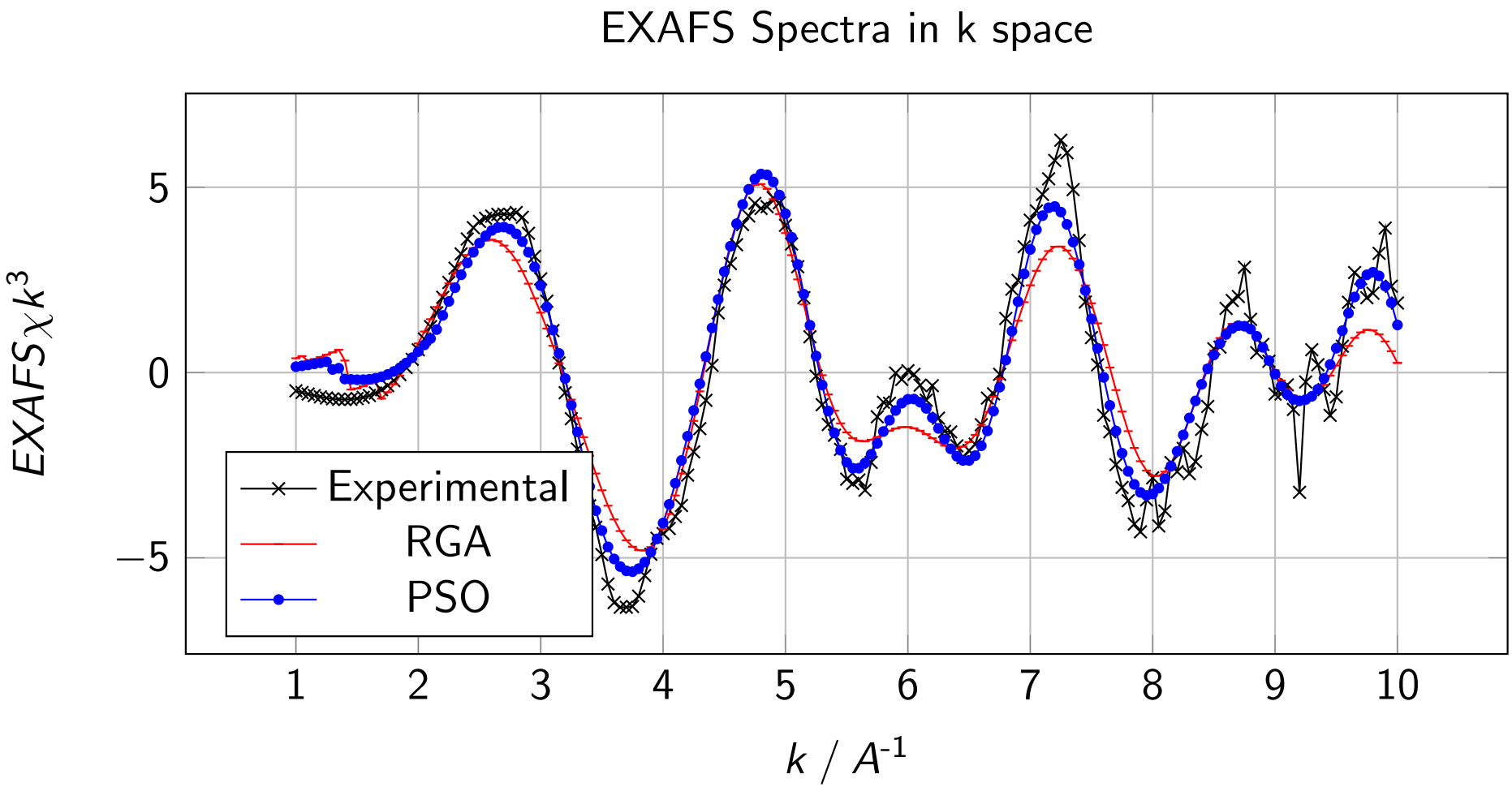


Figure : OEC EXAFS Spectrum Comparison for PSO Post-Optimization

Alternative Algorithms

- PSO performed well as a post-optimization technique.
- Why not compare DE, and PSO directly with GA, and RGA?
- Initial population for DE, and PSO experiments will be from the same pool as the GA, and RGA experiments.

Experimental Setup

Settings	Values
Runs	30
Generations	100, 200
Population Size	50, 100
Starting Velocity	0.01, 0.05, 0.1

Table : System Parameters used in DE and PSO

PSO Results

Pop. Size	Gens.	Start Velocity	Best RMSD	Avg. Best RMSD
50	100	0.01	0.7735	0.9136
50	100	0.05	0.6840	0.9004
50	100	0.1	0.7498	0.9049
50	200	0.01	**0.6109**	0.8025
50	200	0.05	0.6653	0.7907
50	200	0.1	0.6621	0.7933
100	100	0.01	0.7392	0.8775
100	100	0.05	0.6881	0.8856
100	100	0.1	0.7306	0.8836
100	200	0.01	0.6546	0.7750
100	200	0.05	0.6571	**0.7714**
100	200	0.1	0.6688	0.7594

Table : Results for PSO runs

DE Results

Pop. Size	Gens.	Best RMSD	Avg. Best RMSD
50	100	0.9793	1.1120
50	200	**0.9405**	**1.0540**
100	100	1.0357	1.1453
100	200	0.9646	1.0624

Table : Results for DE runs

Analysis

- Population size had no effect on the results.
- DE, and PSO experiments statistically performed better than RGA experiments.
- PSO performed statistically better than DE.
- Might be reaching best approximation due to errors in the EXAFS spectrum.

Best PSO Result

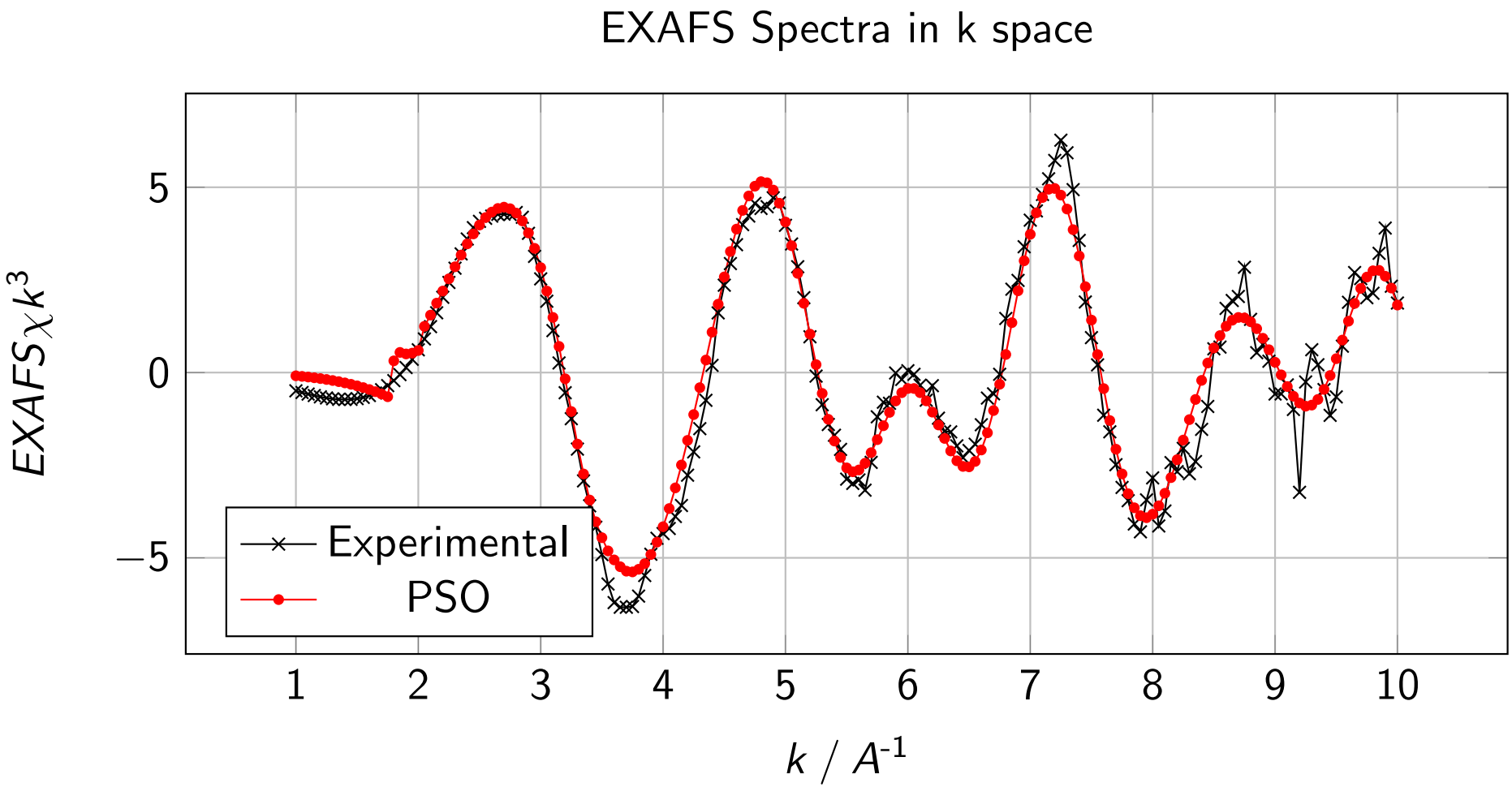


Figure : OEC EXAFS Spectrum Comparison for PSO

Final Comparison

Algorithm	Best RMSD
DFT-QM/MM [1]	1.2679
R-QM/MM [1]	1.2437
GA	1.0533
RGA	0.9649
Post-Optimized PSO	0.7977
DE	0.9405
PSO	**0.6109**

Table : Summary of Best Candidate Solutions

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

Results

Table : Results from subsets experiments

Flexible Atoms	Rigid Atoms	Best	Average
Mn, Ca, C, O, N, H	None	1.1998	1.2580
Mn, Ca, C, O, N	H	**1.1697**	1.2621
Mn, Ca, C	O, N, H	2.4413	N/A
Mn, Ca, O	C, N, H	1.2531	N/A
Mn, Ca, N	C, O, H	2.4881	N/A
Mn, Ca, C, O	N, H	**1.1648**	N/A
Mn, Ca, C, N	O, H	2.5143	2.5649
Mn, Ca, O, N	C, H	N/A	N/A
Mn, Ca	C, O, N, H	2.4916	2.5088

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

Conclusion

- Population based search algorithms performed well on the structure refinement problem.
- Algorithms operating on a continuous search space performed better.
- Biologist needs to verify which candidate solutions are potential solutions.

Future Work

- 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.
- Reduce range on EXAFS spectrum comparison.

Example Spectrum Comparison

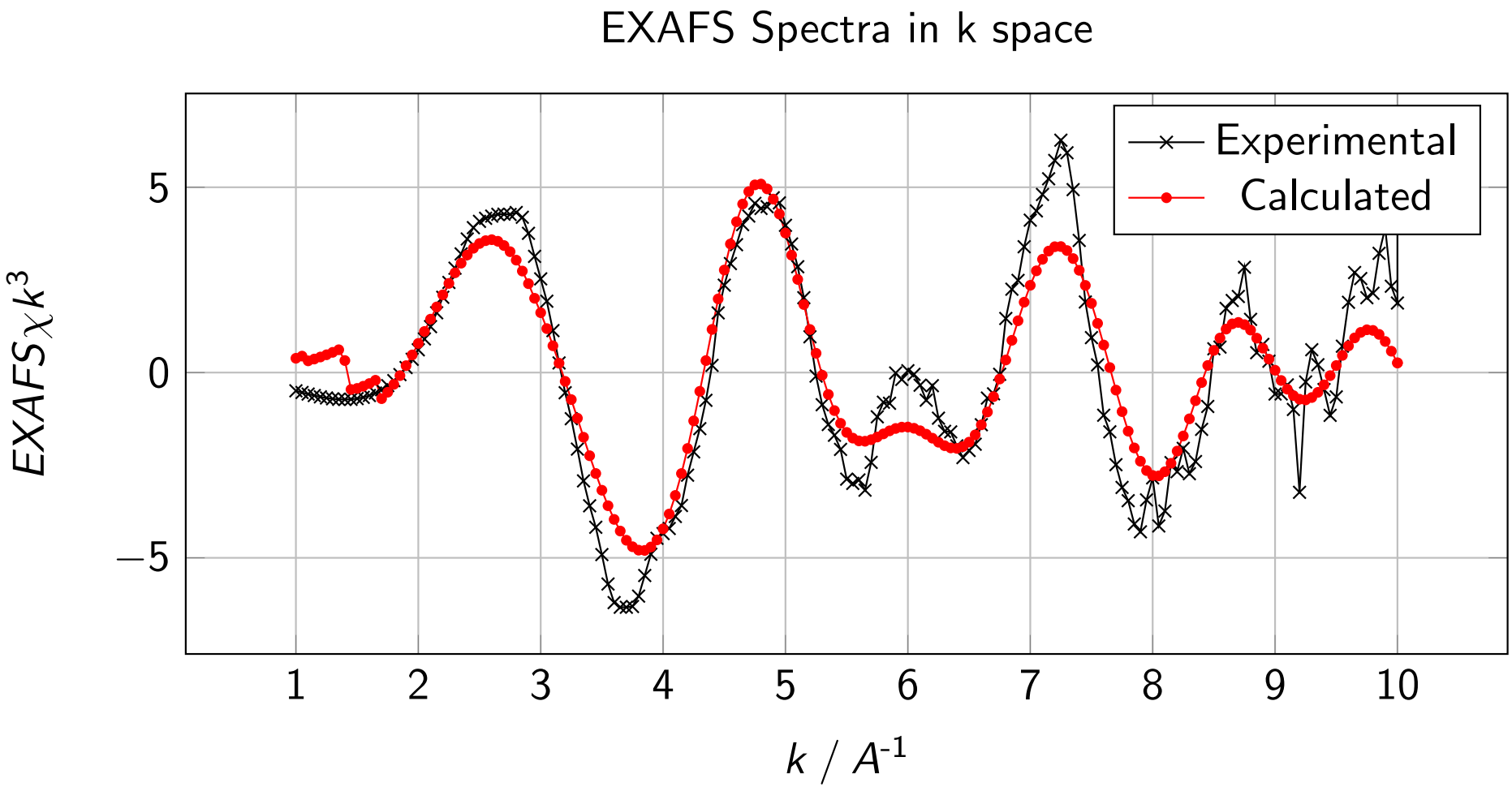


Figure : OEC EXAFS Spectrum Comparison

Thank you.
Questions?

Bibliography

S. Luber, I. Rivalta, Y. Umena, K. Kawakami, J.-R. Shen, N. Kamiya, G. W. Brudvig, and V. S. Batista, "S_1-state model of the O_2-evolving complex of photosystem II," *Biochemistry*, vol. 50, no. 29, pp. 6308–6311, 2011.