Global Optimisation Applied To Molecular Architecture

by

Wayne John Pullan

A thesis submitted toward the degree of

Doctor of Philosophy

Central Queensland University

Department of Mathematics and Computing

Faculty of Applied Science

November, 1996

Abstract

This thesis addresses the problem of identifying configurations of molecular structures which correspond to the globally minimum potential energy for that structure. Molecular structures arise as a result of non-bonded and bonded atomic interactions and experimental evidence shows that, in the great majority of cases, the potential energy global minimum corresponds to the most stable configuration of the molecular structure. This configuration is of particular importance as it dictates most of the physical properties of the molecular structure.

The potential energy of a molecular structure may be calculated, as a function of the atomic positions, using appropriate molecular models. However, as these give rise to potential energy functions that are typically non-convex with many local minima, finding the global minima is an extremely difficult problem. For many years this problem has been investigated by chemists and physicists however, in more recent years, researchers from optimisation and computer science have also become involved and, in fact, the minimisation of non-convex potential energy functions arising from molecular conformation or protein folding problems has become one of the most important interdisciplinary problems [43].

This thesis develops and analyses a molecular structure global optimisation method using both deterministic local and stochastic global optimisation techniques within a genetic algorithm based environment. By incorporating different genetic operators, the one basic method was able to globally optimise a number of different types of molecular structures.

From an experimental point of view, the method was particularly successful and found

- all currently accepted global minima for scaled Lennard-Jones atomic

clusters of 2 to 80 atoms.

- two new global minima for 77 and 78 atom scaled Lennard-Jones atomic clusters.

- all currently accepted and some improved global minima for mixed argon-xenon atomic clusters of 7, 13 and 19 atoms. In addition, minima were determined for all remaining clusters in the $2, \ldots, 20$ atom range.

- all currently accepted global minima for clusters of benzene molecules of 2 to 6 molecules and new minima for clusters of 8 to 12 molecules.

- all currently accepted global minima for a two-dimensional model molecular structure where the number of atoms ranged from 3 to 42.

- currently accepted global minima for a number of small molecules.

Of particular importance is that, in determining these global minima, the method always started from randomly generated initial configurations and at no stage used any heuristic information to accelerate the search.

From a theoretical point of view, this thesis presents an analytical comparison of the phenotype crossover operators used in the method with the more standard (genotype) crossover operators normally used in genetic algorithms. This analysis is confirmed with experimental results. In addition, a proof of convergence for the stochastic global optimisation technique used within the genetic algorithm environment and analytical evaluation of all potential energy gradients required by the deterministic local optimiser are presented.

Chapter 1 of this thesis describes the molecular architecture problem and presents a review of local and global optimisation techniques. Chapter

2 describes the development of APSE, the stochastic global optimisation technique used in this study while the results obtained by applying APSE to the pure atomic cluster problem are presented in Chapter 3.

Chapter 4 describes the development of GEM*, the major computational method used in this study. GEM* implements a combination of local optimisation and APSE probabilistic searches within a genetic algorithm based environment. The results obtained by applying GEM* to the pure atomic cluster problem and a theoretical comparison of phenotype genetic crossover operators with more standard genetic crossover operators are presented in Chapter 5.

The results obtained by applying GEM* to mixed argon-xenon atomic cluster problems are described in Chapter 6 while the optimisation of clusters of benzene and water molecules by GEM* is discussed in Chapter 7. Chapter 8 describes the GEM* optimisation results obtained for a model molecular structure and Chapter 9 presents the GEM* optimisation results for a number of small molecules.

A summary and future research directions are presented in Chapter 10 while the appendices contain the analytical derivation of the potential energy gradients required for the implementation of the BFGS local optimiser and tables describing the structures obtained for mixed atomic clusters.

Within this thesis

- Chapter 2 and Sections 3.3.1 and 3.4.1 appeared in the Australian Computer Journal, Vol. 28, No. 4, November 1996.

- Chapter 6 has been accepted for publication by the Journal of Computational Chemistry.

- Sections 4.2, 5.2, 5.3 and 5.4 have been submitted to the Journal of

Global Optimization.

- Section 3.3.2 appeared as Technical Report 95 – 010, Department of Mathematics and Computing, Central Queensland University.
- Chapter 7 appeared as Technical Report 96 – 005, Department of Mathematics and Computing, Central Queensland University.
- Chapters 8 and 9 appeared as Technical Report 96 – 006, Department of Mathematics and Computing, Central Queensland University.

Contents

List of Figures

List of Tables

Acknowledgments

I first commenced doctoral studies in optimal control theory and, while investigating optimisation theory in that context, had the good fortune to meet my current supervisor, Professor Graham Wood. I certainly must thank him for introducing me to the fascinating field of molecular optimisation. Over the last few years Graham has been very supportive and encouraging. In particular, his high academic standards and professionalism set a standard which I hope that I have gone some way towards achieving. In addition, I must thank him for the provision of Newton, an IBM RS6000 workstation, without which none of this work would have been possible.

During the course of these studies I had the opportunity to be present at a considerable number of seminars/discussion groups on various aspects of global optimisation. I certainly benefited from these occasions and would like to thank Graham, Dr. Victor Korotkich, Professor Zelda Zabinsky (University of Washington) and Dr. Bill Baritompa (University of Canterbury) for sharing their knowledge.

I also owe a considerable debt to the late Professor John Smith who provided me with the opportunity to come to Central Queensland University and commence these studies. Finally, I would like to acknowledge the support I received from Central Queensland University, in particular the Department of Mathematics and Computing and the Faculty of Applied Science, in providing the resources and time away that were so necessary for this work.

Declaration

All computer software used to produce the results contained in this thesis is, with the exception of the BFGS local optimiser, the Powell Direction Set local optimiser and the XMOL software, the work of the candidate.

Except for that attributed to other authors and included for the purpose of setting the appropriate context, all material in this thesis is the work and writing of the candidate. Of the material due to the candidate, only the following has appeared, or is expected to appear, elsewhere:

- Chapter 2 and Sections 3.3.1 and 3.4.1 appeared in the Australian Computer Journal, Vol. 28, No. 4, November 1996.
- Chapter 6 has been accepted for publication by the Journal of Computational Chemistry.
- Sections 4.2, 5.2, 5.3 and 5.4 have been submitted to the Journal of Global Optimization.
- Section 3.3.2 appeared as Technical Report 95 – 010, Department of Mathematics and Computing, Central Queensland University.
- Chapter 7 appeared as Technical Report 96 – 005, Department of Mathematics and Computing, Central Queensland University.
- Chapters 8 and 9 appeared as Technical Report 96 – 006, Department of Mathematics and Computing, Central Queensland University.

Wayne Pullan (C9302023X)