

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, ..., 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

1	Introduction	1
1.1	Overview	1
1.2	Molecular Architecture	2
1.2.1	Atomic Clusters	6
1.2.2	Molecular Clusters	7
1.2.3	Molecules	9
1.3	Local Optimisation	12
1.4	Global Optimisation	13
1.4.1	Random Search	14
1.4.2	Simulated Annealing	14
1.4.3	Genetic Algorithms	15
1.5	Outline of Thesis	20
2	APSE Computational Method	24
2.1	Introduction	24
2.2	APS and APSE Algorithms	25
2.3	Convergence of APSE	29
2.4	Test Functions	32
2.4.1	Beale Function	32
2.4.2	Goldstein-Price Function	34
2.4.3	Rastrigin Function	35

2.4.4	Griewank-2 Function	37
2.4.5	Variable Dimension Test Function	38
2.5	Discussion	40
3	APSE Optimisation of Pure Atomic Clusters	42
3.1	Introduction	42
3.2	Pure Atomic Clusters	43
3.3	Computational Methods	51
3.3.1	APSE and Atomic Build-up	53
3.3.2	APSE and PES Smoothing	54
3.4	Results	65
3.4.1	APSE and Atomic Build-up	65
3.4.2	APSE and PES Smoothing	65
3.5	Discussion	67
4	GEM* Computational Method	70
4.1	Introduction	70
4.2	GEM* Algorithm	72
4.3	GEM* Software	77
4.4	Discussion	78
5	GEM* Optimisation of Pure Atomic Clusters	80
5.1	Introduction	80
5.2	GEM* Computational Method	80
5.3	Results	84
5.4	GEM* Crossover Analysis	94
5.4.1	χ_1 Crossover	98
5.4.2	χ_2 Crossover	100
5.4.3	Theoretical - Numerical Comparison	102

5.5	Discussion	104
6	GEM* Optimisation of Mixed Atomic Clusters	105
6.1	Introduction	105
6.2	Mixed Atomic Clusters	106
6.3	GEM* Computational Method	109
6.4	Results	110
6.5	Discussion	110
7	GEM* Optimisation of Molecular Clusters	121
7.1	Introduction	121
7.2	Molecular Clusters	122
7.2.1	Benzene Clusters	123
7.2.2	Water Clusters	125
7.3	GEM* Computational Method	129
7.3.1	Genetic Encoding	129
7.3.2	Genetic Mutation	131
7.4	Results	131
7.4.1	Benzene Clusters	131
7.4.2	Water Clusters	133
7.5	Discussion	136
8	GEM* Optimisation of a Model Molecular Structure	137
8.1	Introduction	137
8.2	Model Molecular Structure	138
8.3	GEM* Computational Method	141
8.3.1	Genetic Encoding	141
8.3.2	Genetic Crossover	142
8.3.3	Genetic Mutation	143

8.4	Results	143
8.5	Discussion	143
9	GEM* Optimisation of Molecules	147
9.1	Introduction	147
9.2	Molecules	148
9.2.1	United Atom Butane	150
9.2.2	Pseudoethane	150
9.2.3	Acetyl Chloride	152
9.3	GEM* Computational Method	154
9.3.1	Genetic Encoding	154
9.3.2	Genetic Mutation	154
9.4	Results	155
9.4.1	United Atom Butane	155
9.4.2	Pseudoethane	155
9.4.3	Acetyl Chloride	155
9.5	Discussion	155
10	Conclusion	156
10.1	Thesis Summary	156
10.1.1	APSE	157
10.1.2	GEM*	158
10.1.3	Pure Atomic Clusters	161
10.1.4	Mixed Atomic Clusters	161
10.1.5	Benzene Clusters	162
10.1.6	Water Clusters	162
10.1.7	Model Molecular Structure	163
10.1.8	Molecules	163

10.2 Further Work	163
BIBLIOGRAPHY	165
APPENDICES	174
A Analytical Energy Gradients	174
A.1 Pure Atomic Clusters	174
A.2 Mixed Atomic Clusters	175
A.3 Benzene Clusters	175
A.4 Water Clusters	178
A.5 Model Molecular Structure	186
B Argon-Xenon Clusters	189

List of Figures

1.1	Optimal structure - 55 atom Lennard-Jones cluster.	8
1.2	Optimal structure - benzene tetramer.	10
1.3	Octane molecule.	11
1.4	GA parameter mappings to objective function	19
2.1	Beale function - effect of λ and β	33
2.2	Beale function - comparison of APS and APSE.	33
2.3	Goldstein-Price function - effect of λ and β	34
2.4	Goldstein-Price function - comparison of APS and APSE. . .	35
2.5	Rastrigin function - effect of λ and β	36
2.6	Rastrigin function - comparison of APS and APSE.	36
2.7	Griewank-2 function - effect of λ and β	37
2.8	Griewank-2 function - comparison of APS and APSE.	38
2.9	Number of minima for $F(x)$	39
2.10	Comparison of APS and APSE for $F(x)$	40
3.1	Scaled Lennard-Jones pair potential.	43
3.2	Optimal structures - $N = 8, \dots, 13$ atomic clusters.	45
3.3	Lennard-Jones and Morse pair potentials.	54
3.4	Three atom cluster configuration.	56
3.5	Cross-section of PES for $N = 3$	56
3.6	Pair potentials for $v_1(\alpha, r)$	58

3.7	Pair potentials for $v_2(\alpha, r)$	59
3.8	Pair potentials for $v_3(\alpha, r)$	59
3.9	PES around global minimum, $N = 6$, $v_3(\alpha, r)$	60
3.10	PES around global minimum, $N = 18$, $v_3(\alpha, r)$	61
3.11	PES around global minimum, $N = 29$, $v_3(\alpha, r)$	61
3.12	Minimal energy structures - $N = 29, \dots, 31$ atomic clusters. .	69
4.1	GEM* processing context.	77
5.1	GEM* optimised structure - $N = 77$ atomic cluster.	84
5.2	GEM* optimised structure - $N = 78$ atomic cluster.	85
5.3	Minimum, maximum and average population energies, $N = 72$. 85	
5.4	Generations required to optimise atomic clusters, $N = 2, \dots, 80$. 86	
5.5	Multiprocessor scalability of GEM*.	87
5.6	Structures found - GEM* optimisation of an $N = 18$ cluster. 89	
5.7	Structures found - GEM* optimisation of an $N = 18$ cluster. 90	
5.8	Structures found - GEM* optimisation of an $N = 56$ cluster. 91	
5.9	Structures found - GEM* optimisation of an $N = 56$ cluster. 92	
5.10	Structures found - GEM* optimisation of $N = 37, 38$ clusters. 93	
5.11	Initial and final energies for χ_1 crossover.	95
5.12	Initial and final energies for χ_2 crossover.	95
5.13	Energy trajectories for χ_1 crossover.	96
5.14	Energy trajectories for χ_2 crossover.	96
5.15	Effect of close atoms on the BFGS optimised energy.	97
5.16	Expected number of close atoms from a χ_1 crossover.	99
5.17	Expected number of close atoms from a χ_2 crossover.	102
5.18	Expected and average number of close atoms for χ_1	103
5.19	Expected and average number of close atoms for χ_2	103

6.1	Lennard-Jones pair potentials for Ar , Xe	107
6.2	Minimum population energies for $Ar_{10}Xe_{10}$	111
6.3	Optimised structure for Ar_5Xe_2	113
6.4	ΔE for $Ar_{7-n}Xe_n$, $Ar_{13-n}Xe_n$ and $Ar_{19-n}Xe_n$	114
6.5	ΔE for $Ar_{10-n}Xe_n$	114
6.6	GEM* optimised structures - $Ar_5Xe_5, \dots, Ar_2Xe_8$	116
6.7	ΔE for $Ar_{12-n}Xe_n$	117
6.8	GEM* optimised structures - Ar_1Xe_{11}, Ar_0Xe_{12}	117
6.9	ΔE for $Ar_{17-n}Xe_n$	118
6.10	GEM* optimised structures - $Ar_6Xe_{11}, \dots, Ar_3Xe_{14}$	119
7.1	Intermolecular pair potentials for benzene molecules.	124
7.2	Intramolecular pair potentials for water molecules.	128
7.3	Intermolecular pair potentials for water molecules.	128
7.4	Coulombic potentials for water molecules.	129
7.5	GEM* optimised structures - benzene dimer...pentamer. . . .	134
7.6	GEM* optimised structures - benzene hexamer...nonamer. . .	135
7.7	GEM* optimised structures - water dimer and hexamer. . . .	136
8.1	Globally optimal $N = 61$ model molecular structure.	139
8.2	Globally optimal energies - model molecular structure.	139
8.3	Changes in optimal energy - model molecular structure. . . .	140
8.4	Genetic encoding for the model molecular structure.	142
8.5	GEM* optimised structures - model molecular structure. . . .	145
8.6	GEM* optimised structures - model molecular structure. . . .	146
9.1	Structure of the united atom butane molecule.	151
9.2	Structure of the pseudoethane molecule.	151
9.3	Structure of the acetyl chloride molecule.	153

List of Tables

2.1	Summary test results for APS and APSE.	40
3.1	Minimum energies - atomic clusters, $N = 2, \dots, 105$	46
3.2	Counts of Lennard-Jones and Morse local minima.	55
3.3	APSE function evaluations required for $N = 3, \dots, 20$	65
3.4	Energy trajectories for APSE with PES smoothing.	66
3.5	Total energy trajectories for APSE with PES smoothing.	66
6.1	Lennard-Jones parameters for <i>Ar</i> and <i>Xe</i> pair potentials. . .	106
7.1	Structural parameters for benzene.	123
7.2	Potential energy parameters for benzene.	124
7.3	RWK2 potential parameters.	127
7.4	Optimised potential energy values for benzene clusters.	133
7.5	Optimised potential energy values for water clusters.	133
8.1	Performance of search methods - model molecular structure. .	141
8.2	Packed hexagonal energies for the model molecular structure. .	144
9.1	Lennard-Jones parameters - pseudoethane.	152
9.2	Local minima for pseudoethane.	152
9.3	Van der Waals parameters for acetyl chloride.	153
9.4	Acetyl chloride Cartesian coordinates	153

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 Baritempa (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)