

# Contents

Acknowledgements	v
Abstract	vii
Résumé	ix
List of figures	xiii
List of tables	xviii
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation	2
1.2 Overview	3
<b>2 Ab Initio Methods and Molecular Dynamics</b>	<b>5</b>
2.1 Overview	5
2.2 Solving the Time-Independent Electronic Schrödinger Equation	8
2.3 The Hartree-Fock Approximation	9
2.4 Density Functional Theory	10
2.4.1 The Kohn-Sham Equations	12
2.4.2 Exchange-Correlation Functionals	13
2.5 Molecular Dynamics	14
2.5.1 Integrating Equations of Motion: Liouville Formalism	16
2.5.2 Multiple Time Step Integration	18
<b>3 Machine Learning in Chemistry:</b>	
<b>Kernel Methods</b>	<b>21</b>
3.1 Overview	21
3.2 Linear Regression	23
3.3 Feature Maps and Feature Spaces	25
3.4 The Kernel Method	26
3.4.1 Reproducing Kernel Hilbert Spaces	27
3.4.2 Kernel Flavours	29
3.5 Learning From Functions: Non-Linear Regression with Kernels	29
3.6 Learning Potential Energy Surfaces	33

3.6.1	Atomic Decomposition	34
3.6.2	Representing Chemical Environments	34
3.6.3	Kernel Matrix Elements for Learning Potential Energy Surfaces	36
<b>4</b>	<b>Evolutionary and Genetic Algorithms</b>	<b>39</b>
4.1	Overview	39
4.2	Genetic Algorithms	40
4.2.1	Selection Operators	41
4.2.2	Crossover Operators	42
4.2.3	Mutation Operators	45
4.3	Multi-Objective Evolutionary Algorithms (MOEA)	46
4.4	Particle Swarm Optimisation (PSO)	49
4.4.1	Topological Considerations	51
4.4.2	Constriction Particle Swarm Optimisation (CPSO)	51
4.4.3	Fully Informed Particle Swarm Optimisation (FIPSO)	52
<b>5</b>	<b>EVOLVE: An Evolutionary Toolbox for the Design of Peptides and Proteins</b>	<b>53</b>
5.1	Abstract	53
5.2	Introduction	54
5.3	Methods	55
5.3.1	Poly-Alanine Structural Template	55
5.3.2	EVOLVE: Optimisation of Peptidic and Protein Sequences	56
5.4	Results and Discussion	60
5.4.1	Homo-Sequence Exfoliation	60
5.4.2	Sequence Optimisation of the Poly-Alanine $\alpha$ -Helix	62
5.5	Conclusions	68
5.6	Acknowledgements	68
<b>6</b>	<b>Genetic Optimisation of Training Sets for Improved Machine Learning Models of Molecular Properties</b>	<b>69</b>
6.1	Abstract	69
6.2	Introduction	70
6.3	Theory	71
6.4	Results and Discussion	74
6.5	Computational Methods	84
6.6	Supporting Information	85
6.7	Acknowledgments	85
<b>7</b>	<b>Machine Learning Enhanced Multiple Time Step <i>Ab Initio</i> Molecular Dynamics</b>	<b>87</b>
7.1	Abstract	87
7.2	Introduction	88
7.3	Theory	90
7.3.1	<i>Ab Initio</i> Nuclear Forces from Machine Learning	90

7.3.2	Multiple Time Step Molecular Dynamics	92
7.3.3	Implementation	93
<b>7.4</b>	<b>Computational Details</b>	<b>94</b>
7.4.1	Small Water Clusters	95
7.4.2	Liquid Water	95
7.4.3	Training Data: Sample Selection	95
7.4.4	ML Model: Kernel and aSLATM Parameters	96
<b>7.5</b>	<b>Results and Discussion</b>	<b>96</b>
7.5.1	Model Performance	97
7.5.2	Conservation Behaviour and Structural Properties	99
<b>7.6</b>	<b>Conclusion</b>	<b>103</b>
<b>7.7</b>	<b>Acknowledgements</b>	<b>104</b>
<b>8</b>	<b>Conclusions</b>	<b>105</b>
8.1	Summary of Results	105
8.2	Future Directions	107
<b>Bibliography</b>		<b>131</b>
<b>9</b>	<b>Appendix</b>	<b>133</b>
9.1	Supporting Information for EVOLVE: An Evolutionary Toolbox for the Design of Peptides and Proteins	133
9.2	Supporting Information for Genetic Optimisation of Training Sets for Improved Machine Learning Models of Molecular Properties	140
<b>10</b>	<b>Publications</b>	<b>141</b>
<b>11</b>	<b>Curriculum Vitae</b>	<b>143</b>