# TABLE OF CONTENTS

ABSTRACT iv
TABLE OF CONTENTS vi
LIST OF FIGURES xi
LIST OF TABLES xii

Chapter 1. Optimization Strategies for the Design of Protein Sequences and Combinatorial Mutation Libraries

- High-throughput protein engineering 2
- Computational protein design by inverse folding 4
- Beyond single-state inverse folding: multi-state design 6
- Beyond pure computational protein design: library design 13
- Conclusions 15
- References 17

Chapter 2. Dramatic Performance Enhancements for the FASTER Optimization Algorithm

- Abstract 22
- Introduction 23
- Improvements to FASTER 24
  - Original FASTER 24
  - Improvement to starting configurations 25
Chapter 3. An Efficient Algorithm for Multi-State Protein Design Based on FASTER

Abstract 39
Introduction 40
Results and discussion 42

Scoring functions 42
Multi-state Monte Carlo 45
Multi-state FASTER 48
Multi-state iBR 48
Multi-state sPR 49

Rotamer optimization (RO) algorithms 54
Test cases for multi-state design 56
Single-state design problems 56
SSD test cases: MSD-FASTER 57
SSD test cases: MSD-MC 61
Multi-state design of protein G 64
Negative design of calmodulin 68

Conclusions 72
Chapter 4. Development and Validation of Methods for Multi-State Design and Combinatorial Library Design

Abstract 81
Introduction 83
Results and discussion 89
   Designed libraries 89
   Experimental characterization of designed libraries 94
   Origin of destabilizing mutations 101
   Influence of the designed library selection method 103
   The nature of approximation in computational protein design 105
Conclusions 109

Materials and methods 112
   Input structural data 112
   Sequence design specifications and energy functions 112
   Sequence optimization 113
   Combinatorial library design 114
   Library construction, expression, and purification 115
<table>
<thead>
<tr>
<th>Chapter 5.</th>
<th>The Importance of Combinatorial Optimization in the Improvement of Models for Computational Protein Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization in computational protein design</td>
<td>133</td>
</tr>
<tr>
<td>Characteristics of CPD as a tool for protein engineering</td>
<td>136</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appendix I.</th>
<th>Combinatorial Methods for Small Molecule Placement in Computational Enzyme Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>142</td>
</tr>
<tr>
<td>Introduction</td>
<td>143</td>
</tr>
<tr>
<td>Results and discussion</td>
<td></td>
</tr>
<tr>
<td>General calculation procedure</td>
<td>146</td>
</tr>
<tr>
<td>Rotation-translation search</td>
<td>149</td>
</tr>
<tr>
<td>Targeted ligand placement</td>
<td>155</td>
</tr>
<tr>
<td>Sequence design</td>
<td>159</td>
</tr>
<tr>
<td>Conclusions</td>
<td>160</td>
</tr>
<tr>
<td>Methods</td>
<td>161</td>
</tr>
<tr>
<td>Structures and charges</td>
<td>161</td>
</tr>
<tr>
<td>Side-chain rotamer libraries</td>
<td>162</td>
</tr>
</tbody>
</table>
Calculation parameters 165

Energy functions and optimization 166

References 167
LIST OF FIGURES

CHAPTER 3.

1. Graphical depictions of the three MSD sequence selection routines described in the text 47
2. Subroutines used by the MSD sequence selection algorithms 53

CHAPTER 4.

1. The core residues of Gβ1 designed in this study 91
2. The general scheme used to design combinatorial mutation libraries based on computational protein design calculations 92
3. Fraction-unfolded curves derived from the stability determination of library xtal-1 96
4. Fraction-unfolded curves derived from the stability determination of library NMR-1 97
5. Fraction-unfolded curves derived from the stability determination of library NMR-60 98
6. Fraction-unfolded curves derived from the stability determination of library cMD-128 99
7. Each library partitioned into three stability groups 100
8. Correlation between simulation energy and experimental stability for the cMD-128 library 108
9. Detail of the library design method 122
10 Denaturation gradient and elution buffer fluorescence profiles 124
11 Fraction-unfolded profiles between different modes of detection 125
12 Fraction-unfolded profiles between different protein preparations 126

APPENDIX I.
1 Contact geometries specified in small molecule pruning step 147
2 Sample results from test calculations presented in Table 1 150
3 Effect of rotational and translational step sizes 152
4 Targeted placement procedure 155
5 The three clustering moves 163

LIST OF TABLES

CHAPTER 2.
1 Test calculations illustrating performance enhancements for FASTER 30
2 Comparison of the improved FASTER to Monte Carlo 34

CHAPTER 3.
1 Performance of MSD-FASTER when applied to four difficult single-state design problems 60
2 The performance of MSD-MC when applied to four difficult single-state design problems 63
3 Multi-state design of 1GB1, a 60-member NMR ensemble of protein G 67
4 Explicit negative design to increase the binding specificity of calmodulin 71
CHAPTER 4.

1 Combinatorial libraries designed from different sources of structural information 93

APPENDIX I.

1 RMSD and number of wild-type contacts as a function of rotational step size and rotamer library 151
2 RMSD and number of wild-type contacts as a function of rotational and translational step sizes 154
3 Results from targeted placement procedure as a function of rotamer library 158