LIST OF FIGURES

Fig. 1.1 Growth in Medline over the past few years 1
Fig. 1.2 Sugars present in nucleic acids 12
Fig. 1.3 Nitrogenated bases present in DNA 13
Fig. 1.4 A schematic molecular structure view of one DNA strand 14
Fig. 1.5 A schematic molecular structure view of a double strand of DNA 14
Fig. 1.6 A double-stranded DNA sequence represented by strings of letters 15
Fig. 1.7 The nitrogenous bases Adenine, Cytosine, Guanine, and Uracil of RNA 16
Fig. 1.8 The central dogma describes the flow of genetic information in the cell 18
Fig. 1.9 Modeling Process 23
Fig. 1.10 Comparison between Biology and NLP 25
Fig. 3.1 Structural levels - primary, secondary and tertiary structure 62
Fig. 3.2 Components of siRNA secondary structure 65
Fig. 3.3 Helix or stem 66
Fig. 3.4 Internal loops 66
Fig. 3.5 Bulge loops 67
Fig. 3.6 Hairpins 67
Fig. 3.7 Multi-branched loops 68
Fig. 3.8 Inclusions of various grammars 77

Fig. 3.9 details of Chomsky hierarchy 78

Fig. 3.10 Two derivations 81

Fig. 3.11 CFG for the sequence “AGCGAAUCAUGAUGCU” 84

Fig. 3.12 Structure of biological sequence using CFG of Fig. 3.11 84

Fig. 3.13 Modified productions of grammar shown in Fig. 3.11 85

Fig. 3.14 Derivations for the siRNA sequence 86

\[
\text{AGCGUCAGUGACUUGAUGCU}
\]

Fig. 3.15 Derivation tree for the siRNA sequence 87

\[
\text{AGCGUCAGUGACUUGAUGCU}
\]

Fig. 4.1 Parse tree of “ugaaagucg” with equivalent secondary structure 94

Fig. 4.2 Developed Model for siRNA 96

Fig. 4.3 The set of productions P generate siRNA sequence 96

\[
\text{GGACUCAUUGGA}
\]

Fig. 4.4 Parse tree for the siRNA sequence GGACUCAUUGGA 97

Fig. 4.5 Secondary structure of siRNA sequence 99

Fig. 4.6 Inside-Outside algorithm 102

Fig. 4.7 CYK algorithm 105

Fig. 4.8 CYK training algorithm 109

Fig. 4.9 CYK algorithm with structural information 112

Fig. 5.1 Running time of CYK algorithm 121

Fig. 5.2 Running time of the inside-outside algorithm 122

Fig. 5.3 F-measure as function of sensitivity and specificity 126
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 5.4</td>
<td>siRNA sequences</td>
<td>127</td>
</tr>
<tr>
<td>Fig. 5.5</td>
<td>siRNA sequences</td>
<td>128</td>
</tr>
<tr>
<td>Fig. 5.6</td>
<td>siRNA sequences</td>
<td>129</td>
</tr>
<tr>
<td>Fig. 5.7</td>
<td>Distribution of $F_{MN}$ measure</td>
<td>133</td>
</tr>
<tr>
<td>Fig. 5.8</td>
<td>Distribution of $S_{N}$</td>
<td>133</td>
</tr>
<tr>
<td>Fig. 5.9</td>
<td>Distribution of $S_{P}$</td>
<td>134</td>
</tr>
<tr>
<td>Fig. 5.10</td>
<td>Distribution of $F_{MP}$ measure for G2</td>
<td>134</td>
</tr>
<tr>
<td>Fig. 5.11</td>
<td>Distribution of $F_{MN}$ measure for G2</td>
<td>135</td>
</tr>
<tr>
<td>Fig. 5.12</td>
<td>$F_{MN}$ and $F_{MP}$ scores as a function of the confidence parameter</td>
<td>135</td>
</tr>
<tr>
<td>Fig. 5.13</td>
<td>Distribution of $S_{P}$</td>
<td>136</td>
</tr>
<tr>
<td>Fig. 5.14</td>
<td>Distribution of $S_{P}$</td>
<td>136</td>
</tr>
</tbody>
</table>