List of Figures

2.1 An animal cell [29] ...................................................... 7
2.2 Building blocks of DNA ...................................................... 8
2.3 DNA packed within a chromosome [51] ............................... 8
2.4 Central Dogma of Molecular Biology [132] ........................... 10
2.5 The Genetic Code [54] ....................................................... 11
2.6 Combination of two amino acids to produce a di-peptide ...... 12
2.7 A polypeptide chain ....................................................... 13

3.1 Knot from a piece of string .................................................. 18
3.2 Different types of knots- from left to right: the unknot, the trefoil knot, and the figure-eight knot, knots with five crossings [112] ....................................................... 19
3.3 Two types of crossings in knots a) Positive crossing b) Negative crossing ....................................................... 19
3.4 Signing conventions of crossing points in knots ..................... 20
3.5 (a)Twist & (b)Writhe ....................................................... 20
3.6 A loop with linking number 2 ............................................. 21
3.7 Twists & writhes in a link .................................................. 22
3.8 DNA knots [45,124] ....................................................... 25
3.9 Three different types of knots found in proteins and their corresponding structures [143]. ....................................................... 27

4.1 Time domain representation of a noisy signal containing 50 Hz and 120 Hz frequencies ....................................................... 38
4.2 Frequency domain representation of the signal in Figure 4.1 ....................................................... 38
4.3 Typical Fourier spectra for (a) a coding stretch of DNA and (b) a non-coding stretch from S.cerevisiae chromosome III [120] 40
4.5 Power spectrum of Humelafin (D13156) obtained using EIIP sequences [2] ............................... 41
4.6 (a) x(n) (b) y(n) (c) Crosscorrelation of x(n) and y(n) ............... 45
4.7 Crosscorrelation of pyrG gene of M. tuberculosis with a 10-kb region of M. tuberculosis genome showing a peak at the gene region ................................................. 47
4.8 Playing the chaos game with rolls of red, green and blue ........... 48
4.9 First recursion of Sierpinski triangle ........................................ 50
4.10 Second recursion of Sierpinski triangle ................................. 50
4.11 CGR with corners assigned ............................................. 53
4.12 First base T is marked ................................................... 53
4.13 “TA” is plotted ........................................................... 54
4.14 “TAC” is plotted ......................................................... 54
4.15 “TACA” plotted .......................................................... 55
4.16 “TACAG” plotted ......................................................... 55
4.17 Musmusculus (Eukaryote) .............................................. 56
4.18 Homosapiens (Eukaryote) .............................................. 56
4.19 Agrobacterium Tumefasciens (Prokaryote) .......................... 57
4.20 E. Coli K12 (Prokaryote) .............................................. 57
4.21 CGR of a reduced alphabet sequence of a protein (pdb code 2gdm) .............................................. 59
4.22 Spectral plots of hub sequences ...................................... 62
4.23 Spectral plots of nonhub sequences .................................. 62
4.24 Spectral plots of knot sequences ....................................... 62
4.25 Crosscorrelation of a knot protein (pdb code 1pmd) with a synthetic sequence of highly hydrophobic amino acids .................. 68
4.26 Crosscorrelation of a knot protein (pdb code 1pmd) with a synthetic sequence of highly hydrophobic and hydrophobic amino acids .................................................. 68
4.27 Crosscorrelation of a knot protein (pdb code 1pmd) with a synthetic sequence of highly hydrophobic, hydrophobic and neutral amino acids ......................................... 68
4.28 Crosscorrelation of a knot protein (pdb code lpmd) with a synthetic sequence of highly hydrophobic, hydrophobic, neutral and not hydrophobic amino acids ........................................... 69
4.29 Crosscorrelation of a knot protein (pdb code lpmd) with a synthetic sequence of neutral amino acids ......................................................... 69
4.30 Crosscorrelation of a knot protein (pdb code lpmd) with a synthetic sequence of hydrophilic amino acids ........................................... 69
4.31 Correlation area cluster values for knot proteins, membrane proteins and hemoglobin proteins ........................................................... 70
4.32 CGR image of a membrane protein (pdb code 1aij) ......................... 73
4.33 CGR image of a hemoglobin protein (pdb code 1abw) ...................... 73
4.34 CGR image of a keratin protein (pdb code bos621) ......................... 74
4.35 CGR image of a knot protein (pdb code ifqm) .................................. 74
4.36 Division of points in a CGR image ................................................. 75

5.1 Components of a neuron and synapse .................................................. 77
5.2 The mathematical model ....................................................................... 78
5.3 The standard sigmoid function ............................................................... 79
5.4 A perceptron classifier ........................................................................... 80
5.5 Radial basis function network ............................................................... 83
5.6 A multilayer perceptron with one hidden layer ....................................... 85
5.7 Two class linear classification. (The support vectors are indicated with crosses) ................................................................. 90
5.8 Framework of knot predictor .................................................................. 95
5.9 Fourier spectrum of a knot sequence (PDB code -1bcd) ...................... 103
5.10 Fourier spectrum of a unknot sequence (PDB code -1bob) ................. 103
5.11 Crosscorrelation of a synthetic sequence of highly hydrophobic amino acids with (a) a knot sequence (PDB code -1bcd) (b) a unknot sequence (PDB code -1bob) ........................................ 104
5.12 Crosscorrelation of a synthetic sequence of highly hydrophobic and hydrophobic amino acids with (a) a knot sequence (PDB code -1bcd) (b) a unknot sequence (PDB code -1bob) ........ 104

vf
5.13 Crosscorrelation of a synthetic sequence of highly hydrophobic, hydrophobic and neutral amino acids with (a) a knot sequence (PDB code -1bcd) (b) a unknot sequence (PDB code -1bob) .................................................. 105

5.14 Crosscorrelation of a synthetic sequence of highly hydrophobic, hydrophobic, neutral and not hydrophobic amino acids with (a) a knot sequence (PDB code -1bcd) (b) a unknot sequence (PDB code -1bob) .................................................. 105

5.15 Crosscorrelation of a synthetic sequence of neutral amino acids with (a) a knot sequence (PDB code -1bcd) (b) a unknot sequence (PDB code -1bob) .................................................. 106

5.16 Crosscorrelation of a synthetic sequence of hydrophobic amino acids with (a) a knot sequence (PDB code -1bcd) (b) a unknot sequence (PDB code -1bob) .................................................. 106

5.17 CGR of a knot sequence (PDB code -1bcd) ........................................ 107

5.18 CGR of a unknot sequence (PDB code -1bob) ...................................... 107

5.19 ROC graph created for ANN classifier by thresholding a test set. .......................................................... 112

5.20 ROC graph created for SVM classifier by thresholding a test set. .......................................................... 112

B.1 List of 278 knotted entries in PDB .................................................. 123
List of Tables

2.1 List of amino acids ................................................. 10
4.1 EIIP values of nucleotides [2] ................................. 40
4.2 Complex values substituted for nucleotides ................. 46
4.3 Cluster centroids of spectral area values of hub, nonhub and knot proteins ........................................ 63
4.4 Kyte-Doolittle hydrophobicity values of amino acids ....... 65
4.5 Different combinations of hydrophobicity values ............ 65
4.6 Percentage of amino acids chosen for creating synthetic sequences using Table 4.5 ............................... 66
4.7 Percentages of amino acids in knot and membrane proteins 71
4.8 Statistical analysis of HCGRR values of knot and unknot proteins ....................................................... 75

5.1 List of organisms selected for knot dataset ................. 96
5.2 List of organisms selected for unknot dataset ............... 97
5.3 Quantitative values of the 11 features for a sample of knot & unknot protein sequences .............................. 101
5.4 Preprocessed values of the 11 features in Table 5.3 ........ 102
5.5 Classification accuracies obtained with different hidden layers 108
5.6 Results by SVM classifier ........................................... 110
5.7 Performance of classifiers in Homosapiens .................. 113