List of Tables

2.1 Performance of HMM-d, HMM-t, HMM-ModE and HMM-Sub for the sub-family classification of the AGC family of kinases ........ 53

3.1 Coverage (percentage of True Positives identified before the first False Positive) and the average percentage of errors per sequence at the MEP of HMM-d and HMM-ModE for classification of Level-2 sub-families of Class A and Class C GPCR proteins ............... 60

3.2 Application of HMM-d, HMM-t, HMM-ModE and HMM-Sub for function-specific classification of the S/T-Y kinase/atypical kinase/lipid kinase/ATP-grasp fold family ................................................. 76

3.3 Genome-wide enzyme identification for three bacterial genomes (E. coli, B. aphidicola and M. pneumoniae) and one eukaryotic genome (P. falciparum) by ModEnzA and EFICAz ............................... 77

3.4 Conflicting annotations for some EC numbers predicted as belonging to the P. falciparum genome by ModEnzA but not annotated in PlasmoDB ............................................................... 81

3.5 Three-digit annotations for the sequences selected from P. falciparum by TierII ModEnzA profiles ........................................ 83

4.1 Prioritised drug target candidates from M. tuberculosis .......... 103

xiii
4.2 Potential metabolic enzyme drug targets identified from the *P. falciparum* genome .................................................. 107

4.3 Artificial holes (enzymes deliberately knocked off from the *M. tuberculosis* network) and the ranking of the correct genes for each activity .................................................. 112

4.4 Orphan activities (holes) identified from *M. tuberculosis* and the top ranking candidate hypothetical genes for each activity ........ 114