

Appendix section for “Towards Site-based Protein Functional Annotations”

The following graphs shown in this section illustrate the ROC curves from different enzyme families using our proposed methods. The AUC of each method is listed in the main article.

Graph legend: approx=approximate matching method; filter=environment filter method; mrf=MRF motif refinement algorithm; voting=voting with extended motif filter; SVM=Support vector machine with extended motif filter.

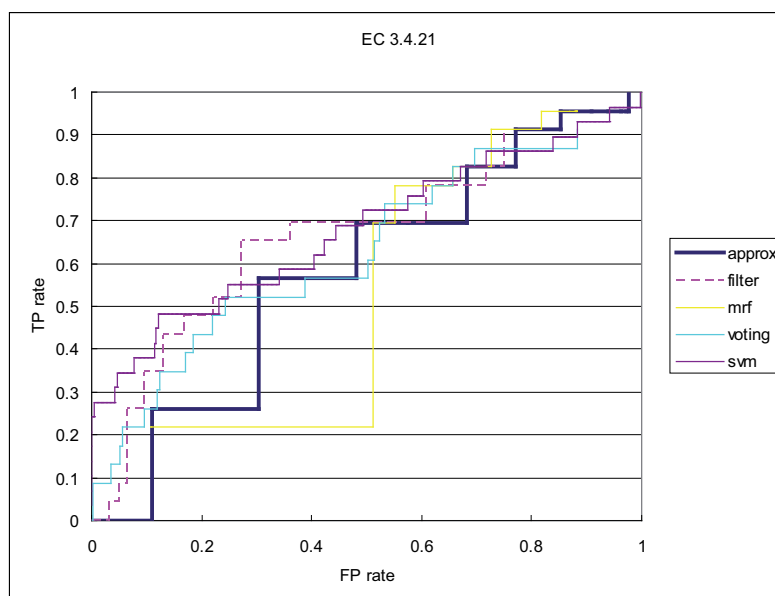


Figure 1: ROC analysis of enzyme family EC 3.2.21.

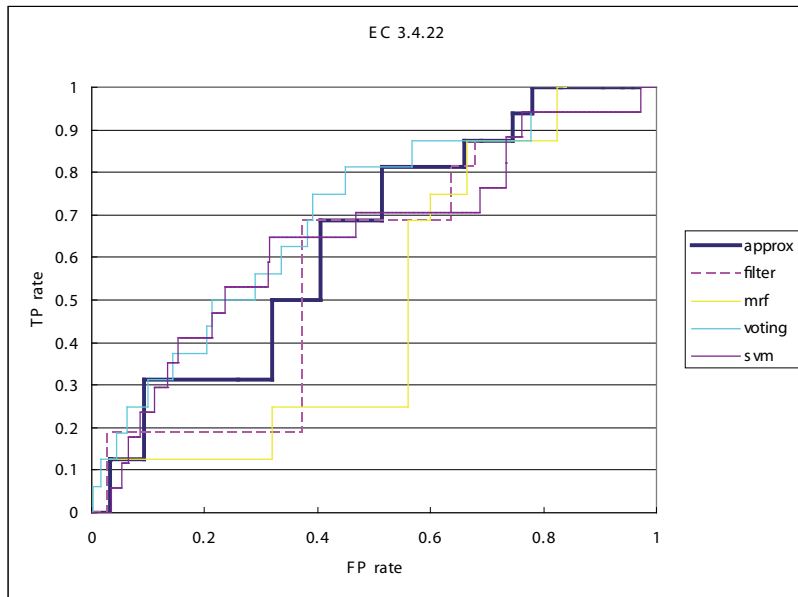


Figure 2: ROC analysis of enzyme family EC 3.2.22.

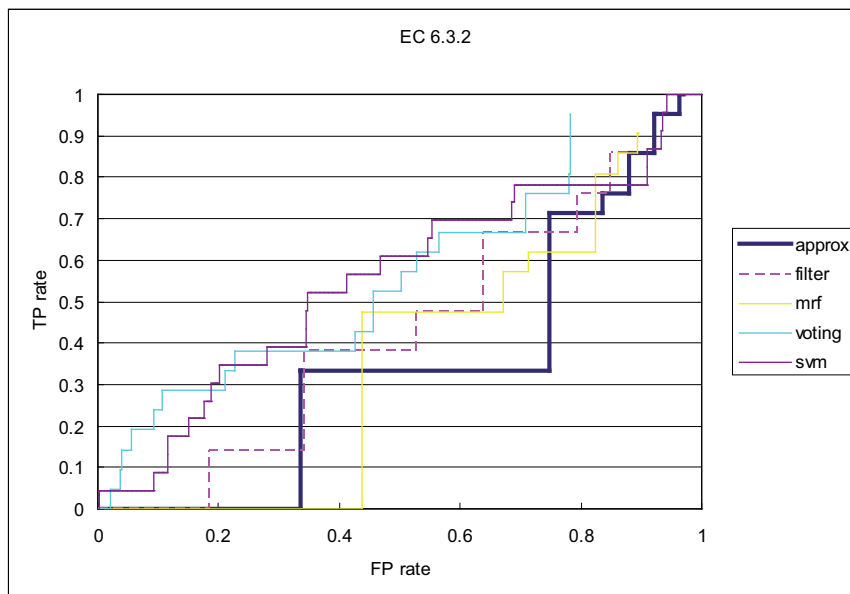


Figure 3: ROC analysis of enzyme family EC 6.3.2.

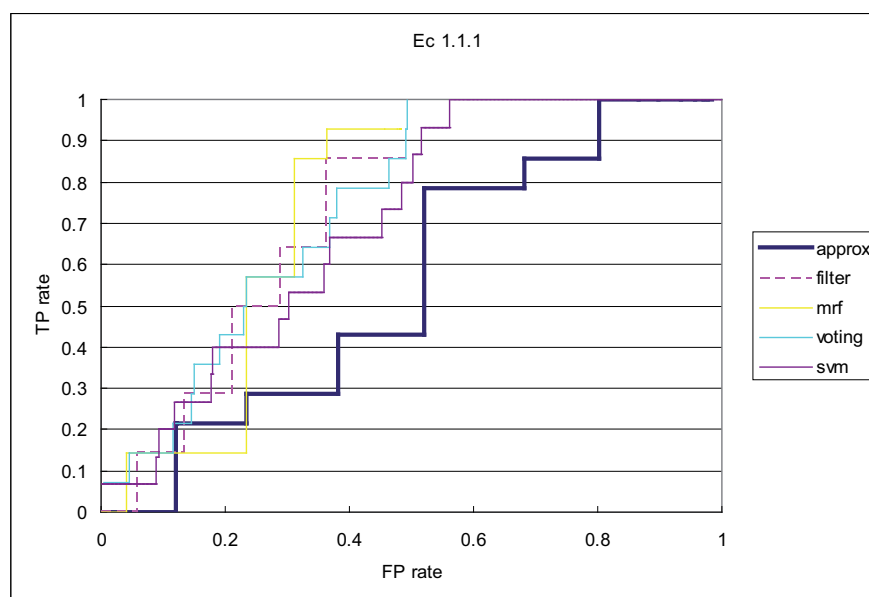


Figure 4: ROC analysis of enzyme family EC 1.1.1.

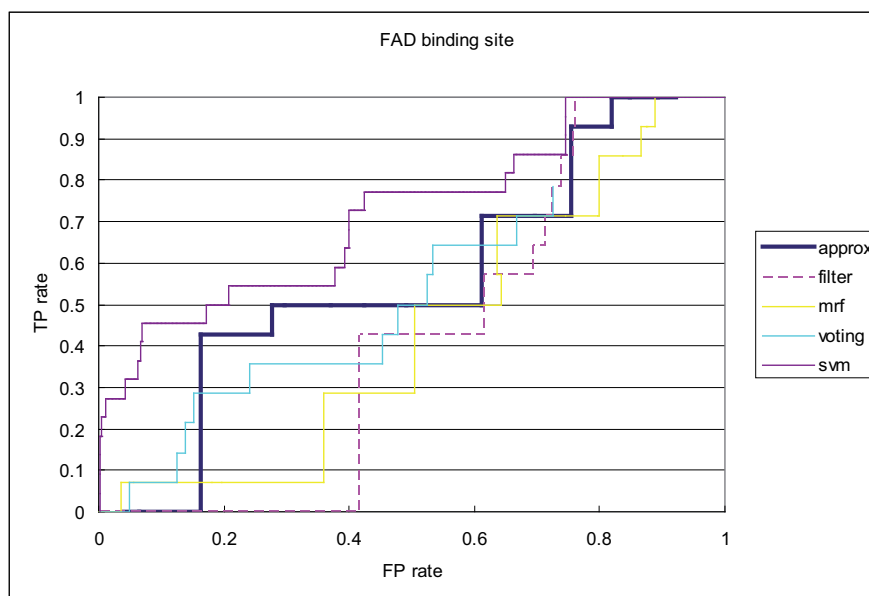


Figure 5: ROC analysis of FAD binding families (1.8.1 + 1.18.1).

EC number	Query protein	Extended motifs
3.4.21	1mcta	ASN95 -HIS57-GLY193-SER195 CYS42 -HIS57-GLY-193-SER195 CYS58 -HIS57-GLY-193-SER195 ILE212 -HIS57-GLY-193-SER195
3.4.22	1pppa	ALA160 -CYS25-HIS159-ASN175 SER29 -CYS25-HIS159-ASN175 VAL157 -CYS25-HIS159-ASN175 ALA27 -CYS25-HIS159-ASN175
6.3.2	2dna	SER19 -GLU15-SER150-GLY276 HIS63 -GLU15-SER150-GLY276 LEU62 -GLU15-SER150-GLY276 THR278 -GLU15-SER150-GLY276
1.1.1	7mdha	LEU200 -ASP201-ARG204-HIS229 VAL265 -ASP201-ARG204-HIS229 GLY227 -ASP201-ARG204-HIS229 ASN173 -ASP201-ARG204-HIS229
FAD binding (1.8.1 + 1.18.1)	1q1ra	HIS43 -GLY11-GLY13-GLY16-ALA20 PRO42 -GLY11-GLY13-GLY16-ALA20 ALA38 -GLY11-GLY13-GLY16-ALA20 GLY111 -GLY11-GLY13-GLY16-ALA20

Table 1: The list of the extended motifs for different EC families, which are used for the voting method and the feature vector method experiments. The bold residues indicate some additional nodes added to the initial motifs.