

## 1 Supplementary Material

While the averaged performance of all methods (structure kernel, FEATURE, GBT) was similar irrespective of the evaluation procedure, we observed that the accuracy was unequal with respect to particular families, superfamilies, and folds (the imbalance was significantly less pronounced with folds). Good classification accuracy of the structure kernel was observed on the following groups: (families) cystathionine synthase-like, arylsulfatase, beta-glycanases, amidinotransferase, and NagB/RpiA/CoA transferase-like; (superfamilies) PLP-dependent transferases, alkaline phosphatase-like, phosphoenolpyruvate/pyruvate domain, pentein, and glutathione synthetase ATP-binding domain-like; (folds) PLP-dependent transferase-like, alkaline phosphatase-like, ATP-grasp, pentein, beta/alpha-propeller and beta-lactamase/transpeptidase-like. Similarly, poor classification accuracy was observed on the following groups: (families) cytochrome c oxidase subunit I-like, phosphoribosyltransferases (PRTases), dUTPase-like, microbial transglutaminase, and LexA-related; (superfamilies) cytochrome c oxidase subunit I-like, metallo-dependent phosphatases, dehydroquinase synthase-like, thiamin diphosphate-binding fold, and LexA/Signal peptidase; (folds) cytochrome c oxidase subunit I-like, metallo-dependent phosphatases, HAD-like, L-aspartase-like and thiamin diphosphate-binding fold. More details about these groups can be found in Supplementary Table 1.

Families	ID	Name	#Sites	<i>AUC</i>
Good	53402	Cystathionine synthase-like	125	99.1
	53653	Arylsulfatase	96	97.9
	51487	beta-glycanases	93	97.4
	55914	Amidino-transferase	68	98.1
	100950	NagB/RpiA/CoA transferase-like	61	99.7
Poor	81441	Cytochrome c oxidase subunit I-like	160	71.7
	53272	Phosphoribosyltransferases	47	67.7
	51284	dUTPase-like	40	65.9
	75336	Microbial transglutaminase	30	60.3
	51307	LexA-related	26	60.4
Superfamilies	ID	Name	#Sites	<i>AUC</i>
Good	53383	PLP-dependent transferases	255	97.6
	53649	Alkaline phosphatase-like	101	97.1
	51621	Phosphoenolpyruvate/pyruvate domain	81	99.4
	55909	Pentain	68	98.0
	56059	Glutathione synthetase ATP-binding domain-like	38	97.8
Poor	81442	Cytochrome c oxidase subunit I-like	160	58.2
	56300	Metallo-dependent phosphatases	60	77.9
	56796	Dehydroquinase synthase-like	54	76.0
	52518	Thiamin diphosphate-binding fold	37	72.0
	51306	LexA/Signal peptidase	26	59.9
Folds	ID	Name	#Sites	<i>AUC</i>
Good	53382	PLP-dependent transferase-like	255	97.6
	53648	Alkaline phosphatase-like	101	97.3
	56058	ATP-grasp	89	96.7
	55908	Pentain, beta/alpha-propeller	68	97.9
	56600	beta-lactamase/transpeptidase-like	43	98.3
Poor	81443	Cytochrome c oxidase subunit I-like	160	71.2
	56299	Metallo-dependent phosphatases	80	76.7
	56783	HAD-like	67	79.4
	48556	L-aspartase-like	56	66.8
	52517	Thiamin diphosphate-binding fold	37	73.2

Table 1: The protein families/superfamilies/folds with which Structure Kernel performs well and poorly