

Supplementary Table 2. Mean accuracy of the k-means clustering using energy terms for classes E, F and G.

SCOP class		E	F	G
All <i>ETs</i>	<i>ACC</i>	0.65	0.54	0.61
	<i>SPC</i>	0.66	0.49	0.55
	<i>SN</i>	0.71	0.60	0.59
	<i>MCC</i>	0.32	0.08	0.23
	<i>F1</i>	0.66	0.56	0.61
<i>ETs</i> different for > 60% domains	<i>ACC</i>	0.65	-	0.61
	<i>SPC</i>	0.65	-	0.63
	<i>SN</i>	0.70	-	0.61
	<i>MCC</i>	0.30	-	0.23
	<i>F1</i>	0.63	-	0.59
2 most differentiating <i>ETs</i>	<i>ACC</i>	0.68*	0.60	0.60
	<i>SPC</i>	0.76*	0.59	0.53
	<i>SN</i>	0.71*	0.62	0.70
	<i>MCC</i>	0.34*	0.19	0.23
	<i>F1</i>	0.65*	0.60	0.62
1 most differentiating <i>ETs</i>	<i>ACC</i>	0.68*	0.59	0.61
	<i>SPC</i>	0.76*	0.56	0.63
	<i>SN</i>	0.71*	0.64	0.61
	<i>MCC</i>	0.34*	0.34	0.22
	<i>F1</i>	0.65*	0.64	0.59
1 <i>ET</i> : total	<i>ACC</i>	0.66	0.54	0.61
	<i>SPC</i>	0.65	0.44	0.53
	<i>SN</i>	0.71	0.62	0.71
	<i>MCC</i>	0.30	0.30	0.23
	<i>F1</i>	0.63	0.63	0.63
3 <i>ETs</i> : <i>hack_elec</i> , <i>p_aa_p</i> , <i>rama</i>	<i>ACC</i>	0.70	0.59	0.60
	<i>SPC</i>	0.84	0.59	0.62
	<i>SN</i>	0.68	0.60	0.61
	<i>MCC</i>	0.44	0.19	0.22
	<i>F1</i>	0.68	0.59	0.59

The models were divided into 2 classes. Only domains with at least 3 mirror and 3 native models were included. Squared Euclidean distance was used. ACC denotes accuracy (Eq. 5), SPC denotes specificity (Eq. 6), SN denotes sensitivity (Eq. 7), MCC denotes Matthews correlation coefficient (Eq. 8), and F1 denotes F1 score (Eq. 9).

* The results of clustering in class E for 2 most differentiating *ETs* and 1 most differentiating *ETs* are the same, because the 4 *ETs* had the same values of the domains for which were significantly different. For that reason they were taken all to cluster as 2 most differentiating *ETs* and 1 most differentiating *ETs*.