Supplementary Information

Blind Dataset #2 Performances. Both the Support Vector Machine-based and the GOR method-based classifiers were tested on two randomized, non-redundant blind datasets (main text, Section 3.1, Blind Datasets). We evaluated the performance of protein secondary structure prediction with the canonical scoring indexes and the Segment Overlap score (main text, Section 3.4). Performance results in the blind tests were similar and concordant to those observed in the 5-fold cross-validations performed for both predictors on the training data (main text, Section 3.1). We reported here the performance results obtained from the Blind Dataset #2 (Table 1), while the assessment of the blind test performed on the Blind Dataset #1 is displayed in the main text.

Method availability . Our Support Vector Machine-based secondary structure classifier is available as a serialized Python object for reuse at `https://github.com/alussana/SVM-II-Str-classifier` `/blob/master/dataset/svm/rbf_2.0_0.3_model.pkl.gz`

Table 1. Blind Test Performance

Blind Test Dataset #2		GOR	SVM
PPV	H	65.19	84.71
	E	53.95	80.47
	C	74.06	67.02
TPR	H	75.86	74.84
	E	69.77	61.18
	C	49.36	84.22
MCC	H	0.513	0.692
	E	0.467	0.623
	C	0.424	0.561
SOV	H	59.04 ± 2.09	63.49 ± 2.43
	E	67.89 ± 1.87	64.64 ± 2.04
	C	49.47 ± 1.39	68.36 ± 1.25
Q_3		63.96	75.19
SOV_3		60.22 ± 1.04	67.19 ± 1.24