

Figure S1: Comparison of characteristic signals of continuous (red circles) and discontinuous (black crosses) DNA molecules in the crystal. The (a) largest local intensity average and (b) top rotation score are plotted against the cube root of the reciprocal unit cell volume.

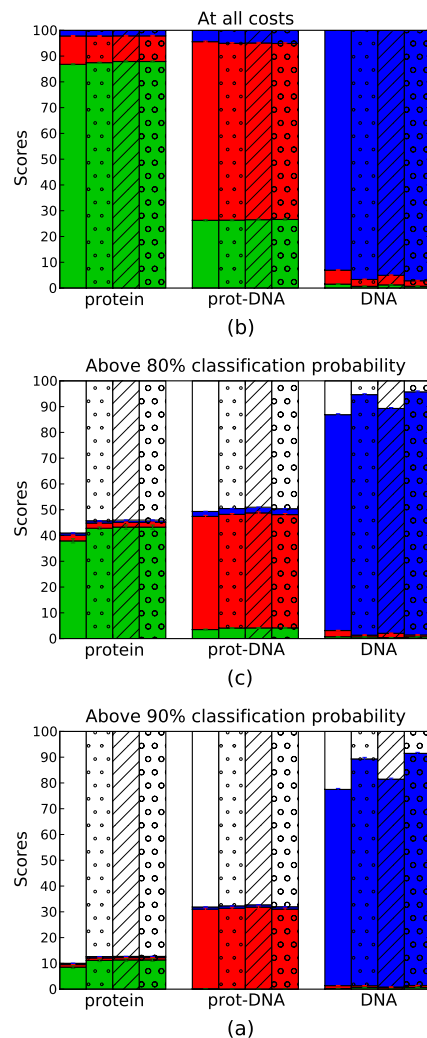
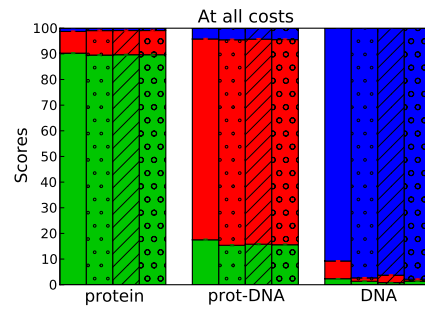
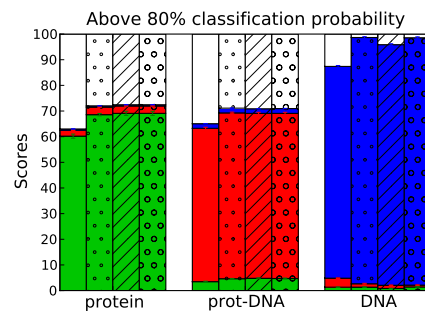


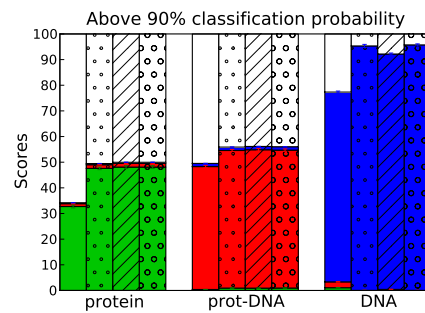
Figure S2: Standalone mode classifier performance for different training and test sets. Plain bars illustrate the results for the control set with all structures. Dotted bars describe the results after removing structures with DNA, but with less than two base pairs in B-DNA conformation. Hatched bars show the results for the set after removing structures with pseudorigin peaks (threshold 40% of the origin peak). Bars with open circles summarize the results after removing structures with pseudorigins or very short B-DNA fragments. Colour coding is like in Fig. 9 (green for protein only, red for protein and DNA, blue for DNA only and white for no prediction). Panels (a), (b) and (c) are for classification at all costs and with greater than 80% and 90% correct classification probability, respectively.



(b)

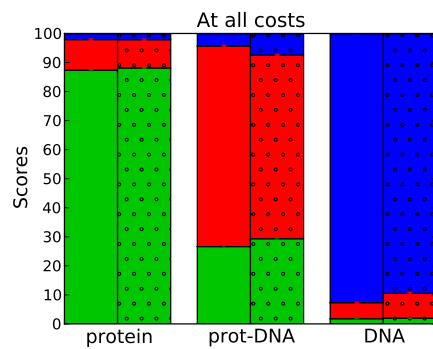


(c)

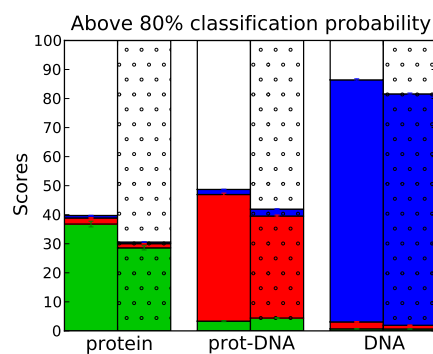


(a)

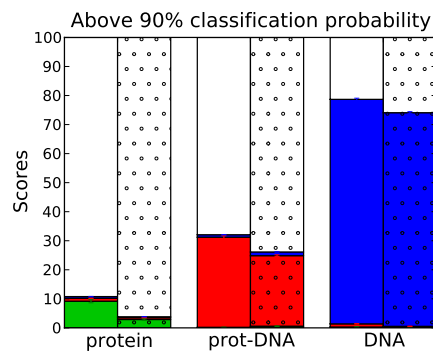
Figure S3: Combined mode classifier performance for different training and test sets. The same symbols and colours as in Fig. S2 are used.



(a)

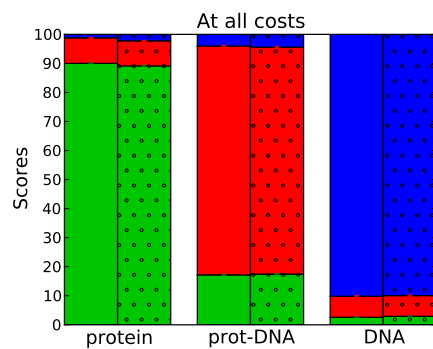


(b)

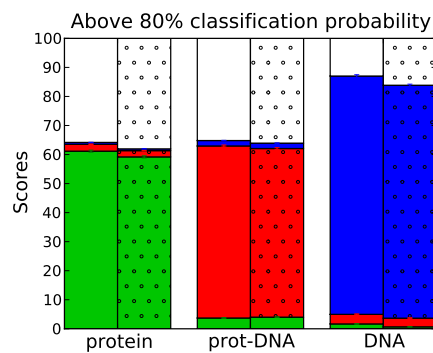


(c)

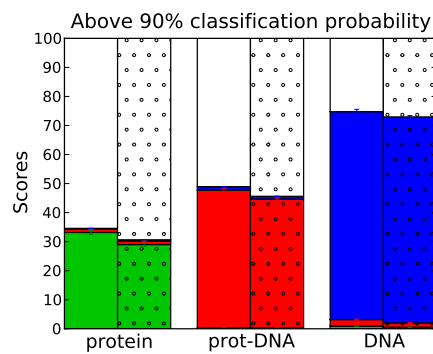
Figure S4: Standalone mode classifier performance with different measures of unit cell size. The cube root of the reciprocal unit cell volume (plain bars) is compared with the inverse of the smallest unit cell dimension (dotted bars). Colours and panels are as in Fig. S2.



(a)



(b)



(c)

Figure S5: Combined mode classifier performance with different measures of unit cell size. The same symbols and colours as in Fig. S4 are used.

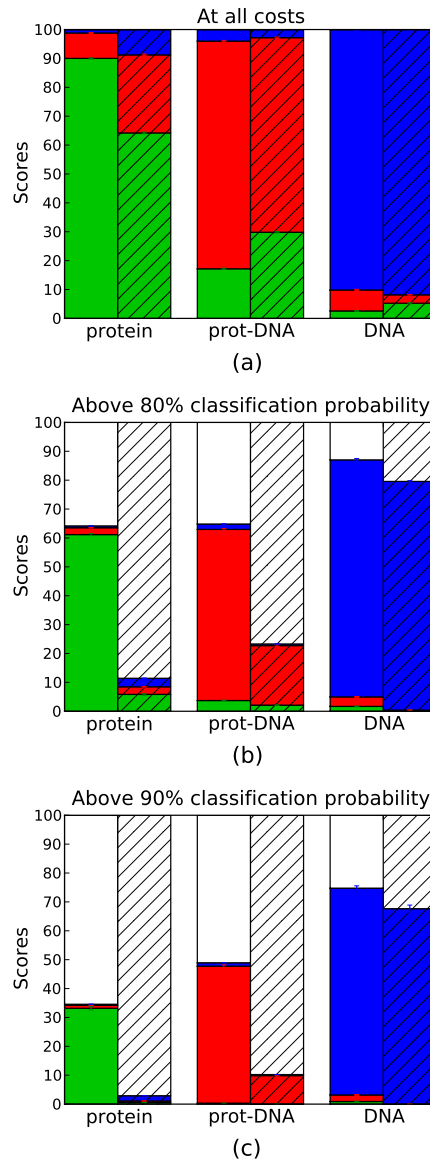


Figure S6: The rotation score (plain bars) versus Z-score (hatched bars) as an input parameter for the classifier in *Phaser* only mode. Both scores were calculated with the same *Phaser* settings and combined with the standard cube root of the reciprocal unit cell volume for classification. Colour coding is like in Fig. S2.