Supplementary material

Sequence-Similar, Structure-Dissimilar Protein Pairs in the PDB

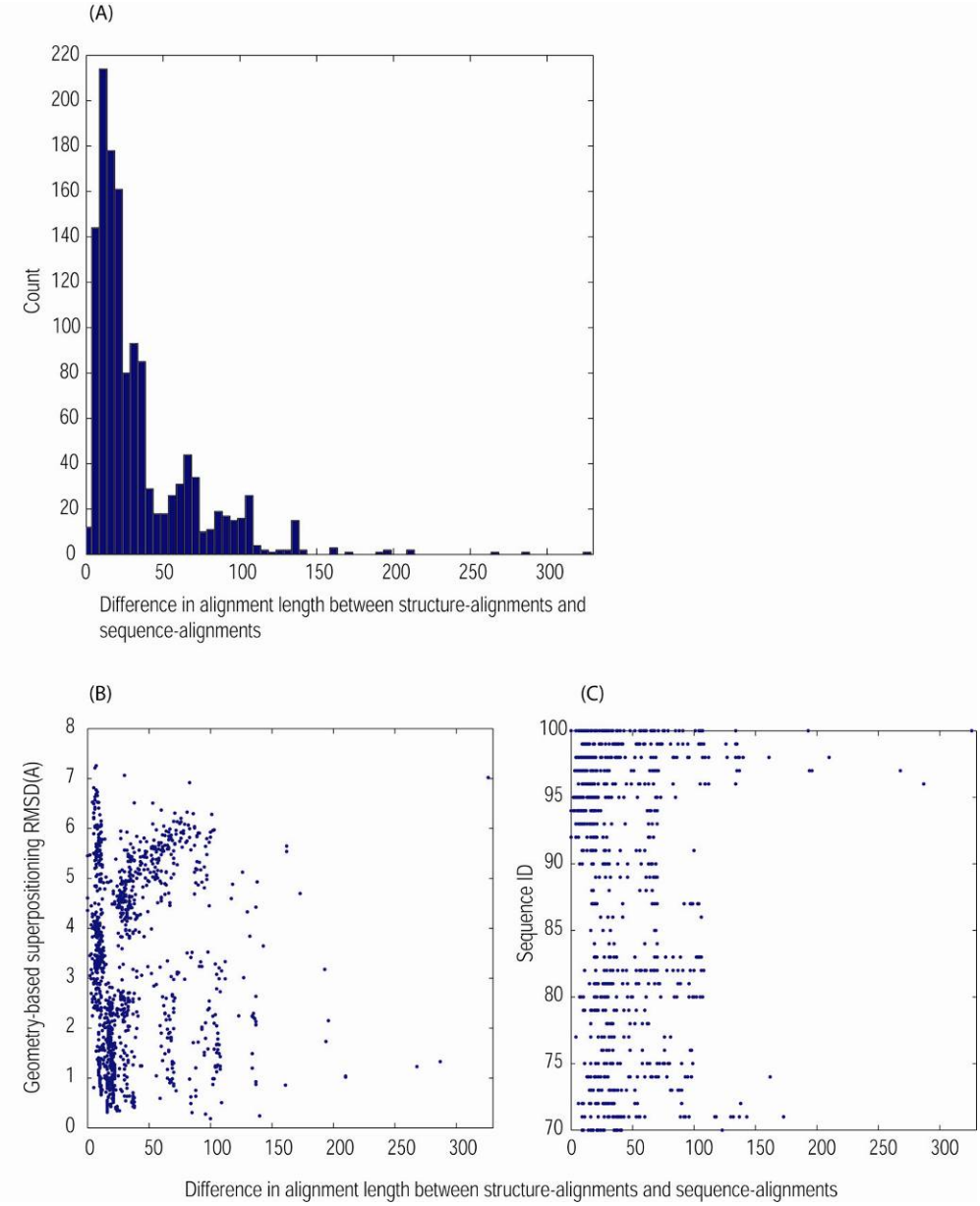

Figure S1: Differences between geometry-based alignments and sequence-alignments in a subset of pairs with sequence-based structure-superpositioning RMSD≥6.
The data shown in this figure is of a subset of pairs with at least 70% sequence identity, as in Figure 2.
(A) Histogram of length differences between sequence alignments and structure (geometry-based) alignments.
(B) Structure-alignment RMSD as a function of the length differences between sequence alignments and structure alignments. Notice that in many pairs with a small difference in the length of the alignments, structure alignments measures an RMSD≥3.
(C) Sequence identity as a function of the length differences between sequence alignments and structure alignments.

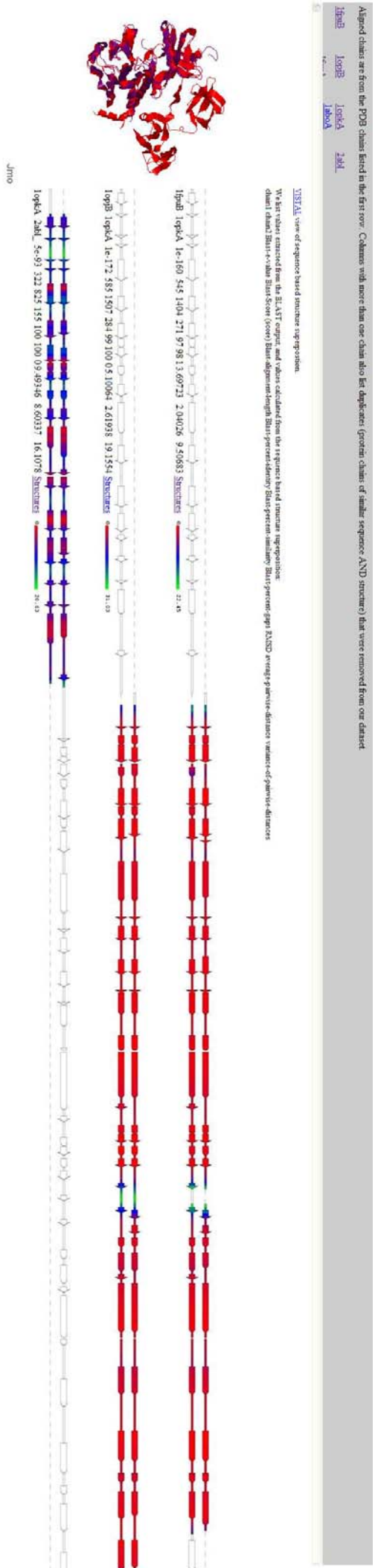

Figure S2: Screen shot of the database entry for the alignment of the structurally-dissimilar entries for the cABL kinase.
1opkA contains both the N-terminal SH2-SH3 domains and the C-terminal kinase domain of the cABL kinase. Both of these parts of cABL can adopt more than one conformation, as can be seen in the alignment of the N-terminal domains of 1opkA to 2abl and in the alignments of the C-terminal domain of 1opkA to 1fpuB and 1opjB. Note that the difference between the SH2-SH3 domains is mainly an inter-domain difference, while the dissimilarity in the kinase domain is intra-domain.