



**Fig. S7.** Representative sequences of known structure from the MAFFT alignment of 2,193 SH2 domains. This corresponds to the same sequences and domain footprint as the GISMO alignment in Fig. S6. In order to emphasize the conserved core, columns containing more than 50% deletions in the complete alignment were removed; these correspond to the insert regions shown here.