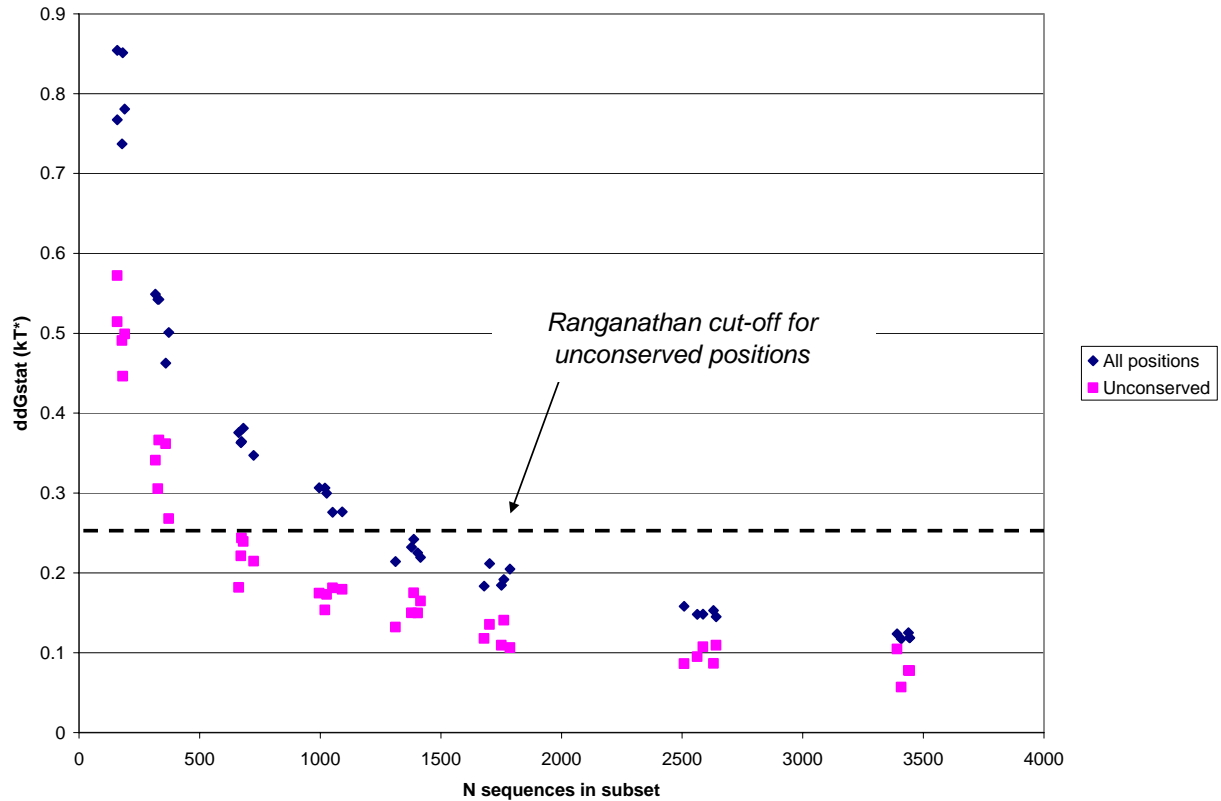


## Supplementary Materials

Magliery &amp; Regan

“Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif”



**Supplemental Fig. 1. Significance of subset sizes.** In order to estimate the statistical significance of the subset size for coupling calculations, we calculated the  $\Delta\Delta G_{stat}$  values for randomly selected subsets of various sizes. Ideally, for each position, the amino acid distribution of the subset will be the same as the distribution in all sequences. If so, the  $\Delta\Delta G_{stat}$  values would be nil. Above, the average  $\Delta\Delta G_{stat}$  values over all 34 positions (blue diamonds) or over the 5 least conserved positions (i.e., lowest  $\Delta G_{stat}$ , pink squares) are plotted as a function of the number of randomly selected sequences. Hatley et. al have suggested the indicated average  $\Delta\Delta G_{stat}$  level for unconserved residues as a cut-off for statistical significance.<sup>1</sup> (Actually, they suggest  $0.025 kT^*$  for scaling to  $N=100$ , but here we use  $N=1000$ . We have shown that  $\ln P$ , and therefore  $\Delta G_{stat}$ , is proportional to  $N$  for large  $N$ .) Note that subsets with about 700 sequences fall under the cutoff (we used 10% or 688). However, it is not clear to us why this particular value is especially useful. The average  $\Delta\Delta G_{stat}$  in all positions is less than  $0.4 kT^*$  for 688 sequences, which is much less than the  $2.5 kT^*$  value we chose to indicate a significant redistribution. There are 6,887 TPR sequences in the full data set.

1. Hatley, M.E., Lockless, S.W., Gibson, S.K., Gilman, A.G. & Ranganathan, R. (2003). Allosteric determinants in guanine nucleotide-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 14445-14450.