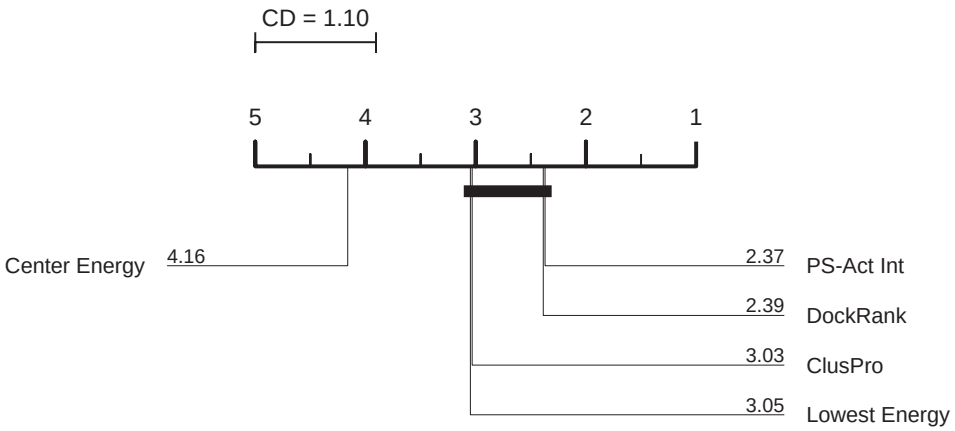

Figure S6. Pair-wise comparisons of different docking scoring methods on the ClusPro2-BM3_31 decoys using the Nemenyi test. Methods that are not significantly different (at significance level $\alpha = 0.05$) are grouped together (via connecting lines). The number on the line of each method is the average rank of each method over docking cases. The pairwise Nemenyi test shows that the average L-RMSDs of top models selected by DockRank are significantly smaller than those selected by ClusPro Center Energies. However, the average L-RMSDs of top models selected by ClusPro, Lowest Energy and DockRank are not significantly different, which indicates that DockRank has limited improvement on the highest ranked models in terms of L-RMSDs when applied to the docked models by pre-filtered ClusPro scoring functions. The definition of a hit is a docked conformation with $L\text{-}RMSD \leq 10$ Å.