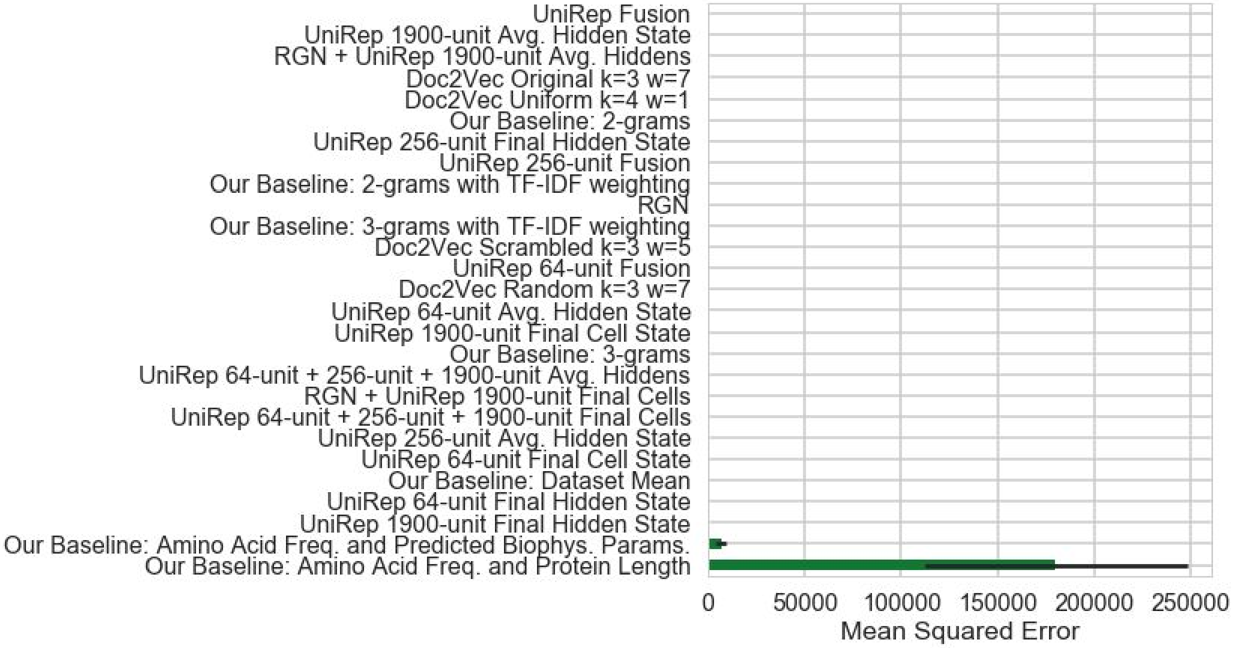
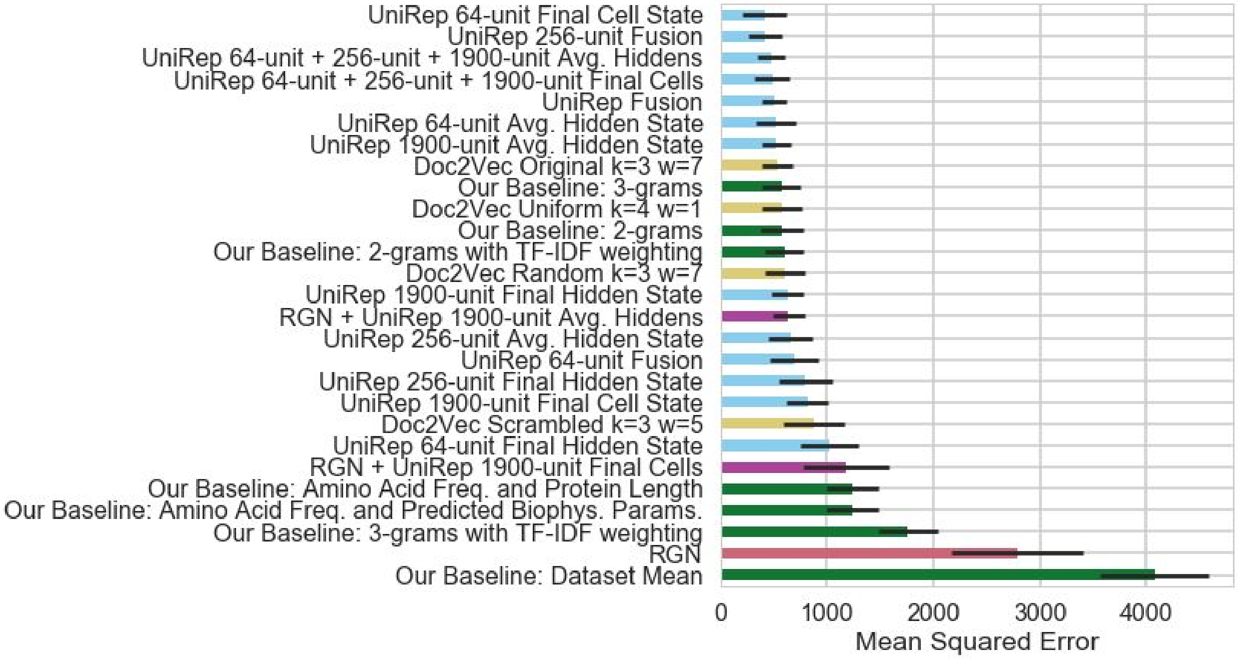
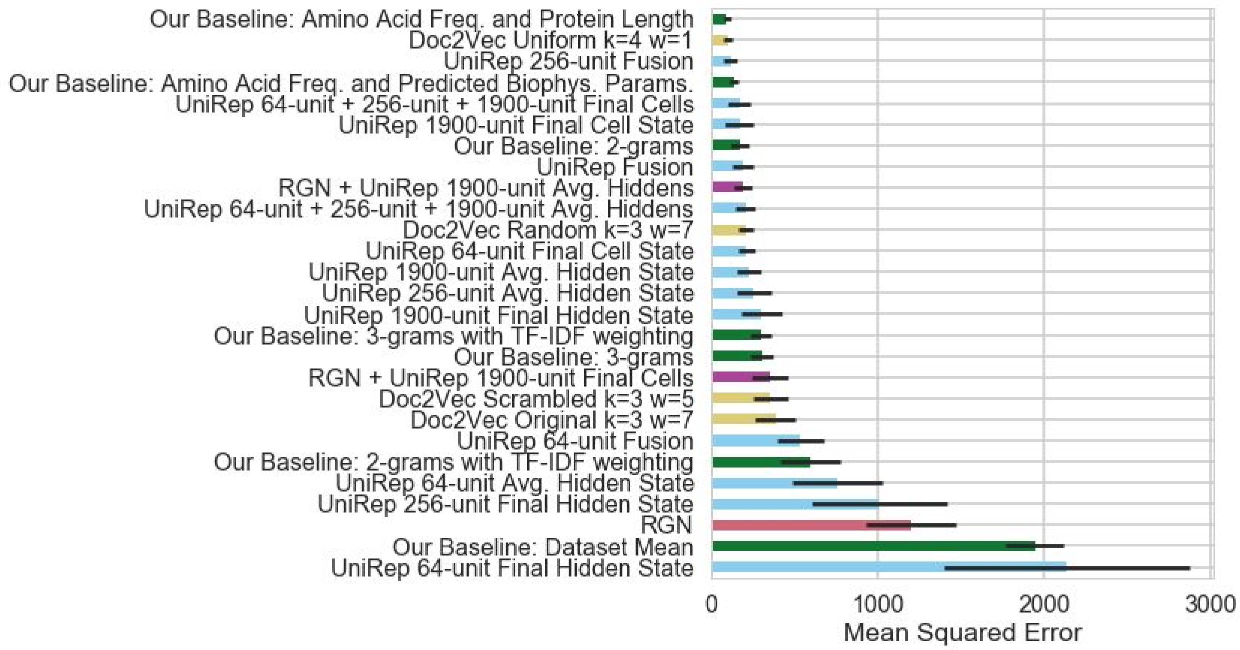


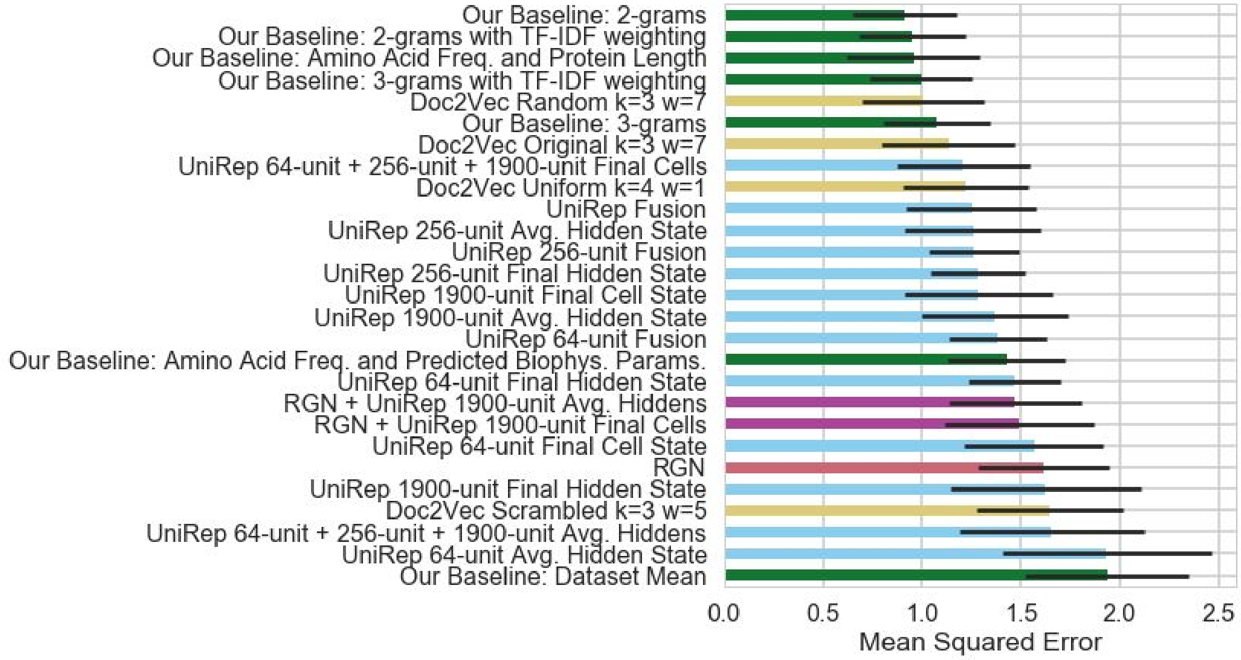
Supplementary Results Figure 1: All representation results in cytochrome p450 thermostability task.



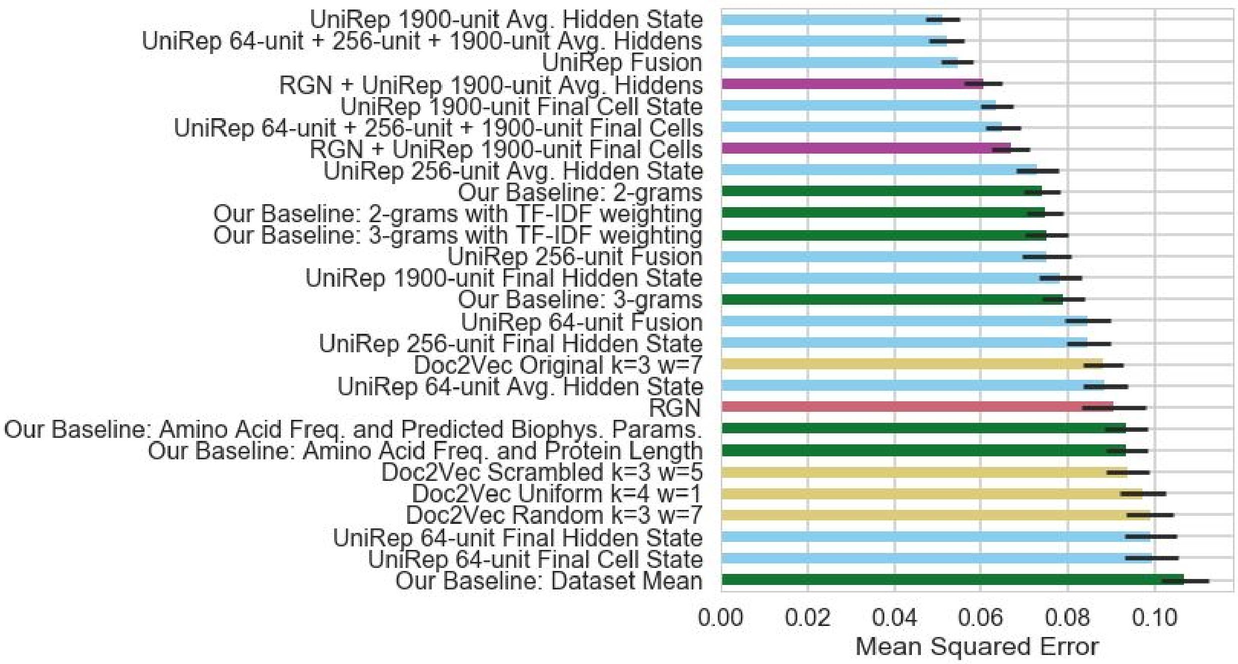
Supplementary Results Figure 2: All representation results in rhodopsin peak absorption wavelength task.



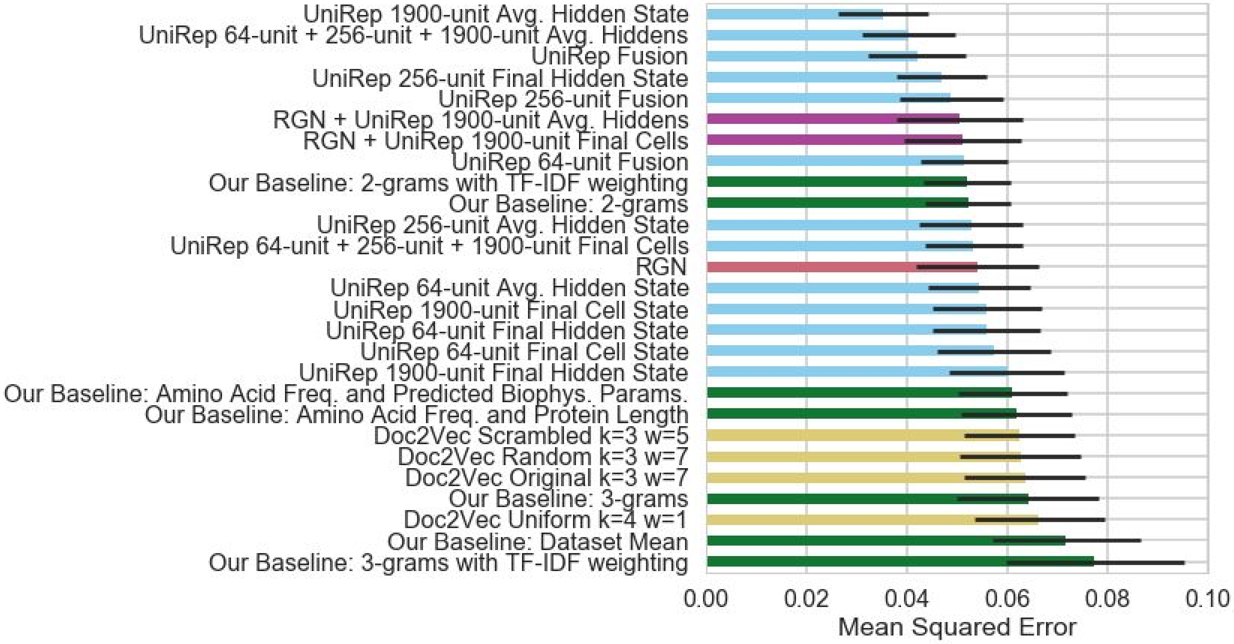
Supplementary Results Figure 3: All representation results in epoxide hydrolase enantioselectivity task.



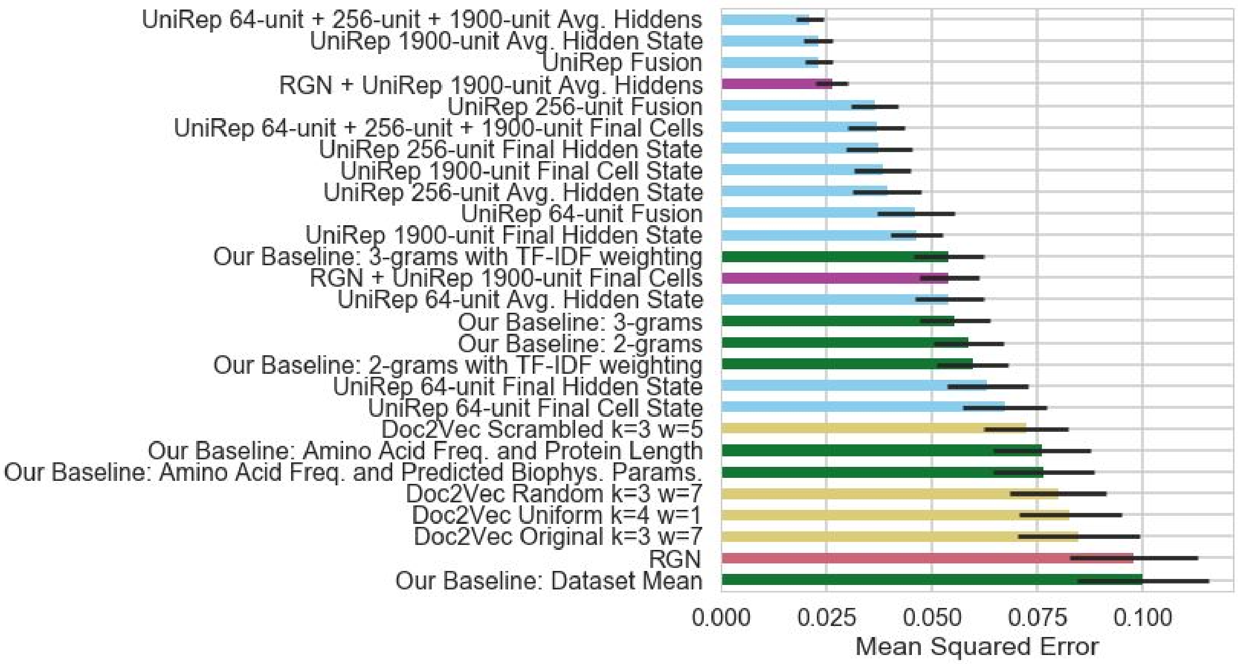
Supplementary Results Figure 4: All representation results in channelrhodopsin membrane localization task.



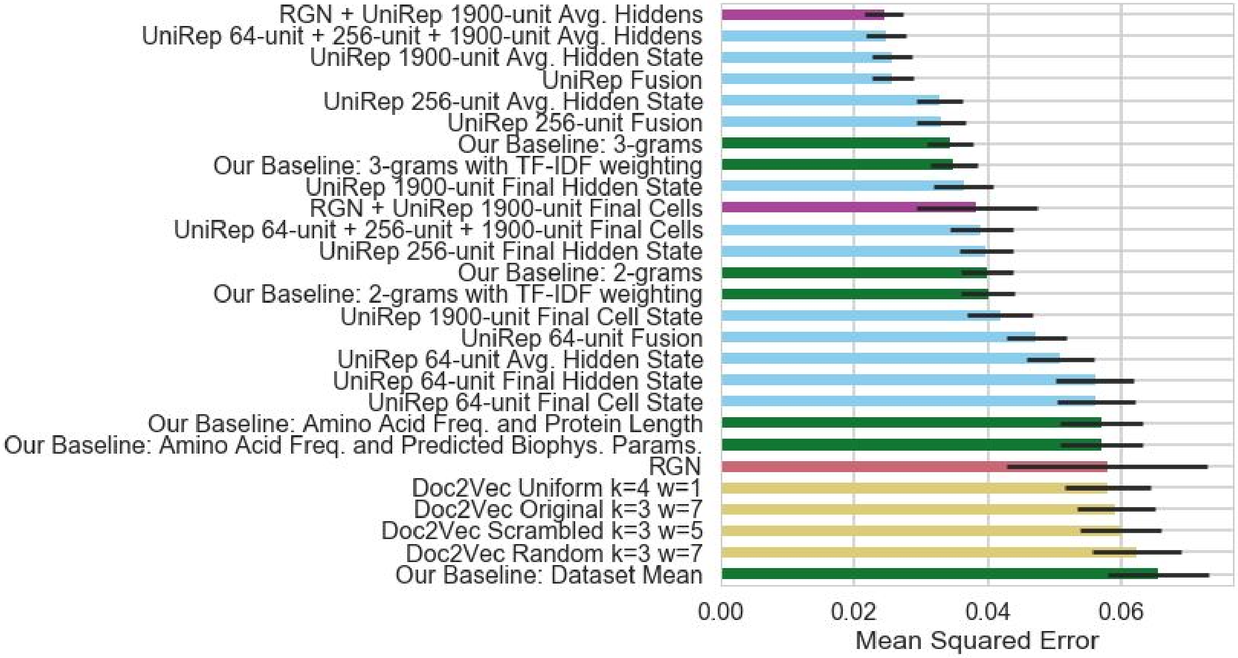
Supplementary Results Figure 5: All representation results in variant effect prediction task task. Subset: TEM-1 Beta-lactamase.



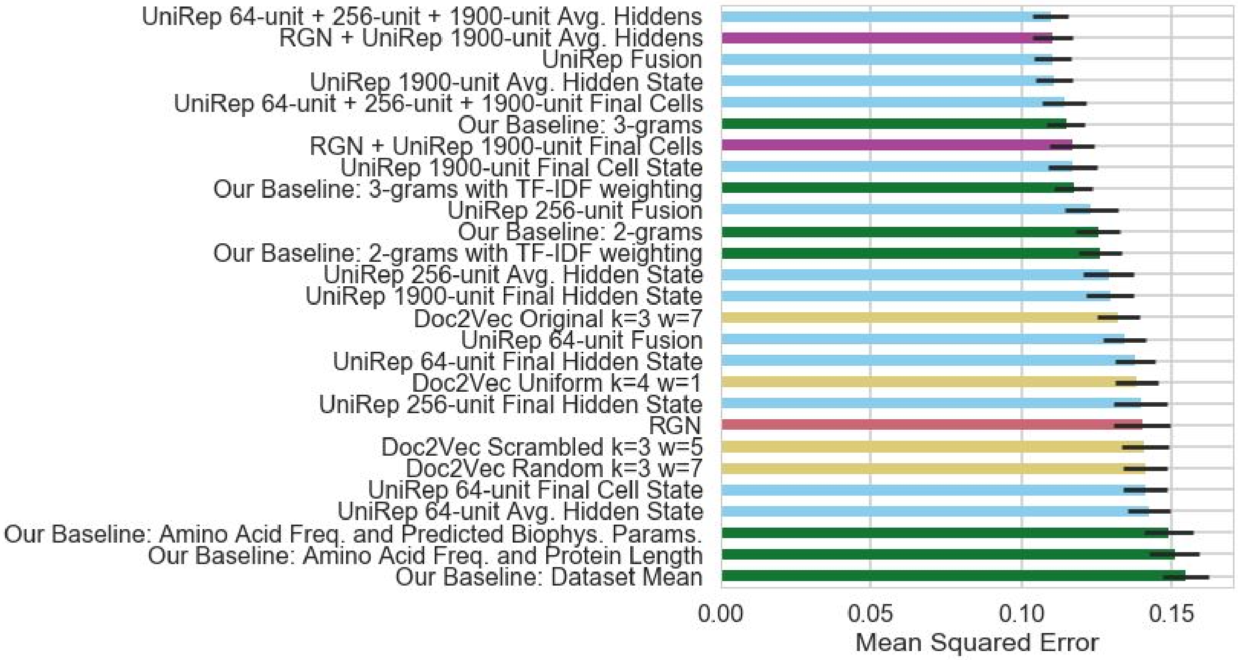
Supplementary Results Figure 6: All representation results in variant effect prediction task task. Subset: Ubiquitin (E1 Activity).



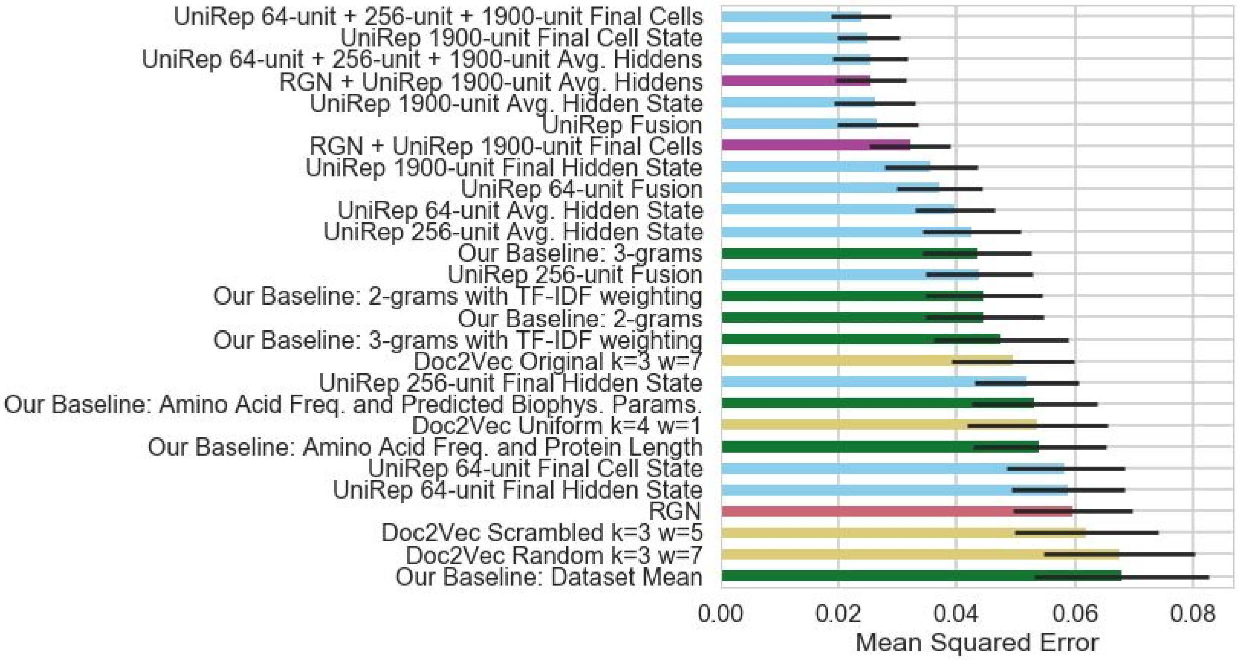
Supplementary Results Figure 7: All representation results in variant effect prediction task task. Subset: Protein G (IgG domain).



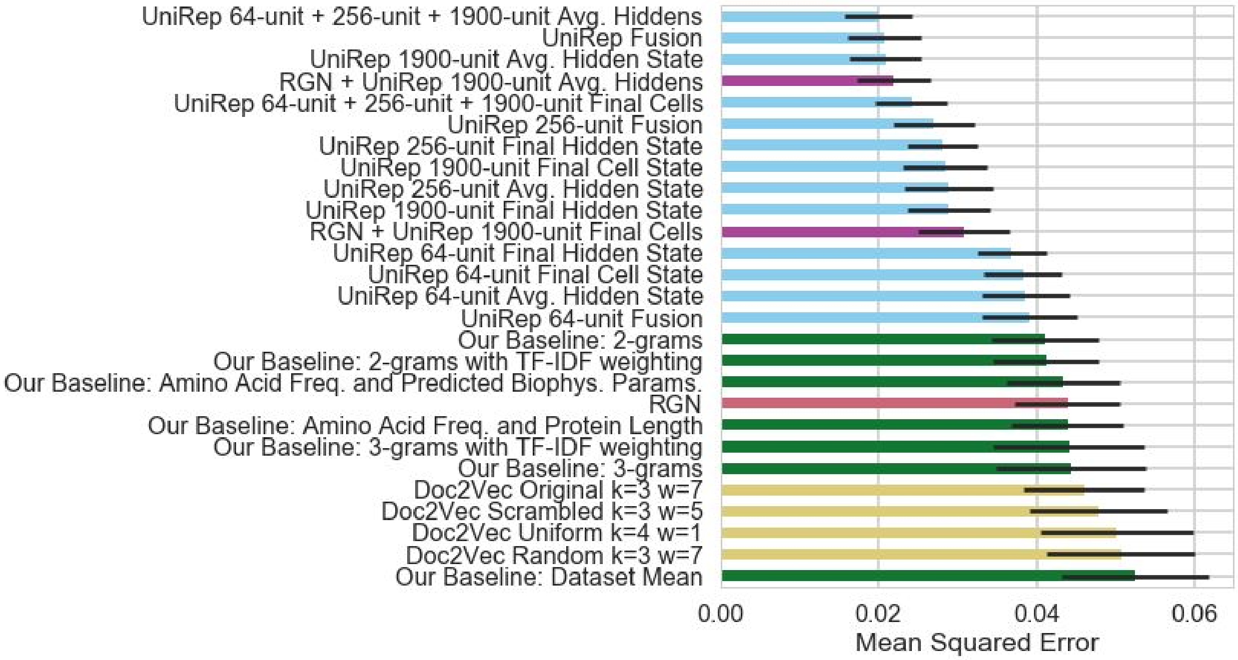
Supplementary Results Figure 8: All representation results in variant effect prediction task task. Subset: HSP90.



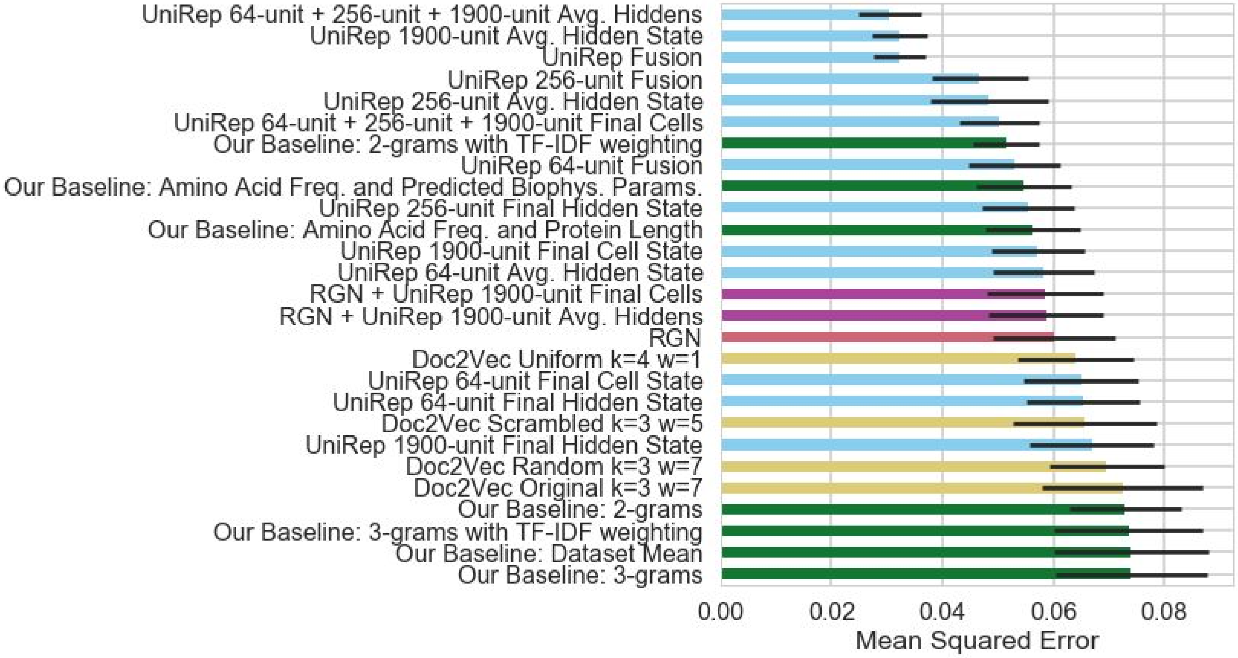
Supplementary Results Figure 9: All representation results in variant effect prediction task task. Subset: Aminoglycosidase (Kka2).



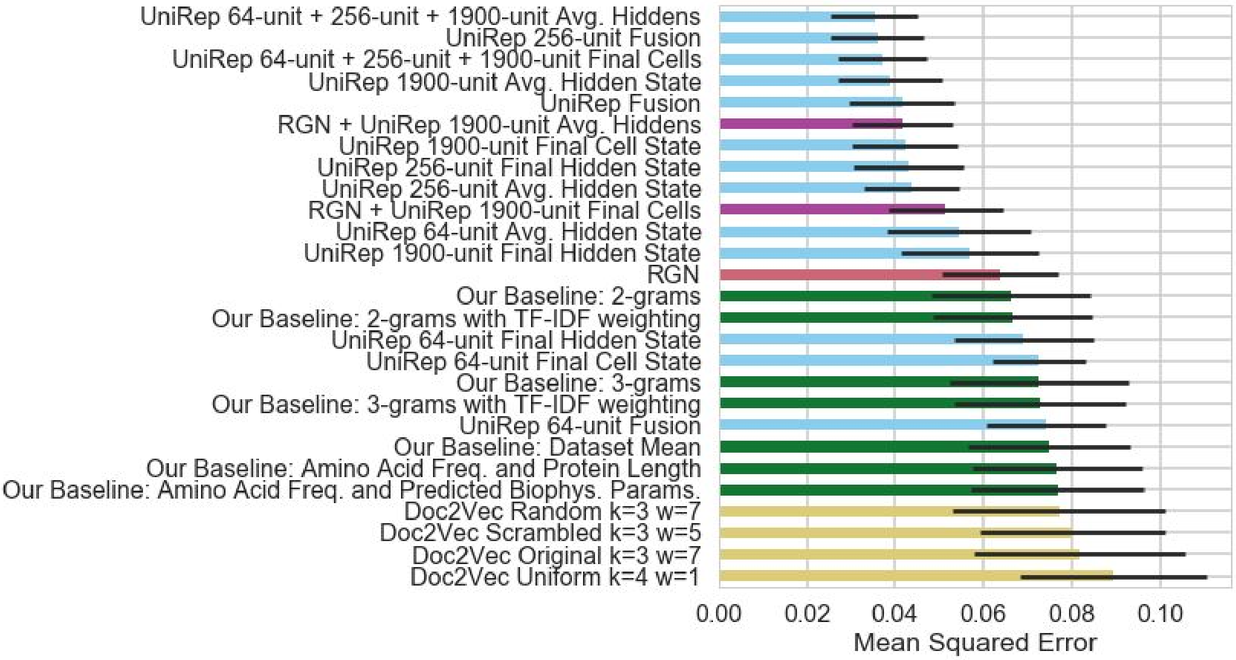
Supplementary Results Figure 10: All representation results in variant effect prediction task task. Subset: Pab1 (RRM domain).



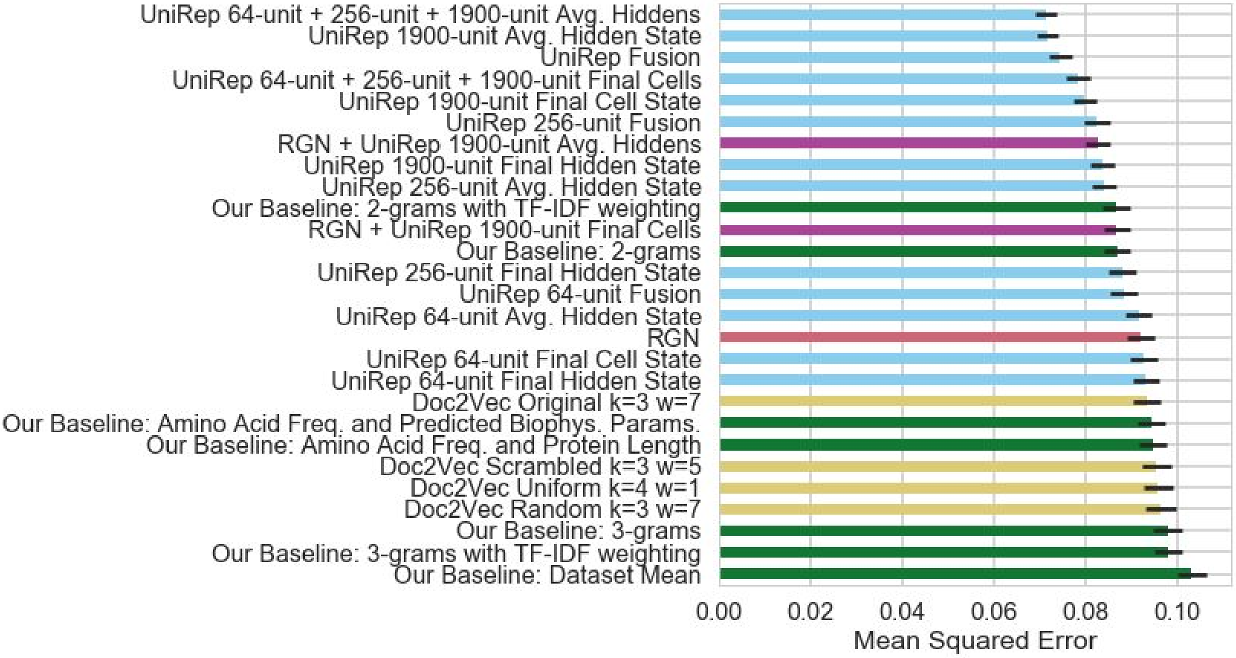
Supplementary Results Figure 11: All representation results in variant effect prediction task task. Subset: PSD95 (Pdz3 domain).



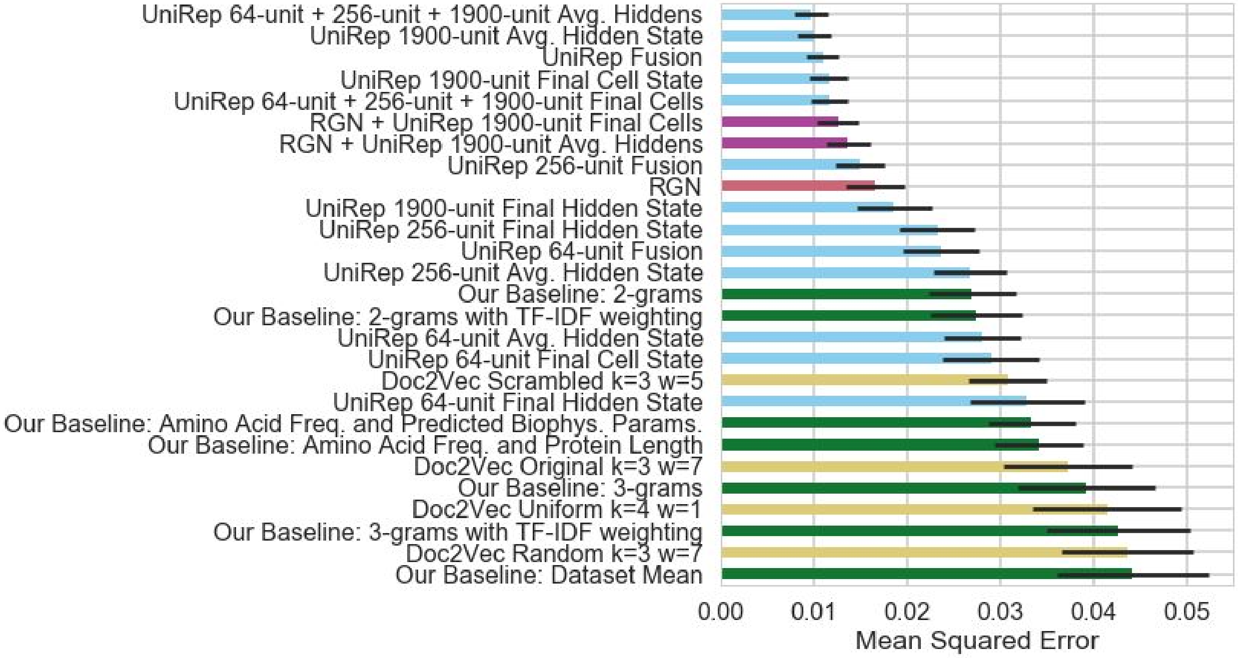
Supplementary Results Figure 12: All representation results in variant effect prediction task task. Subset: Ubiquitin.



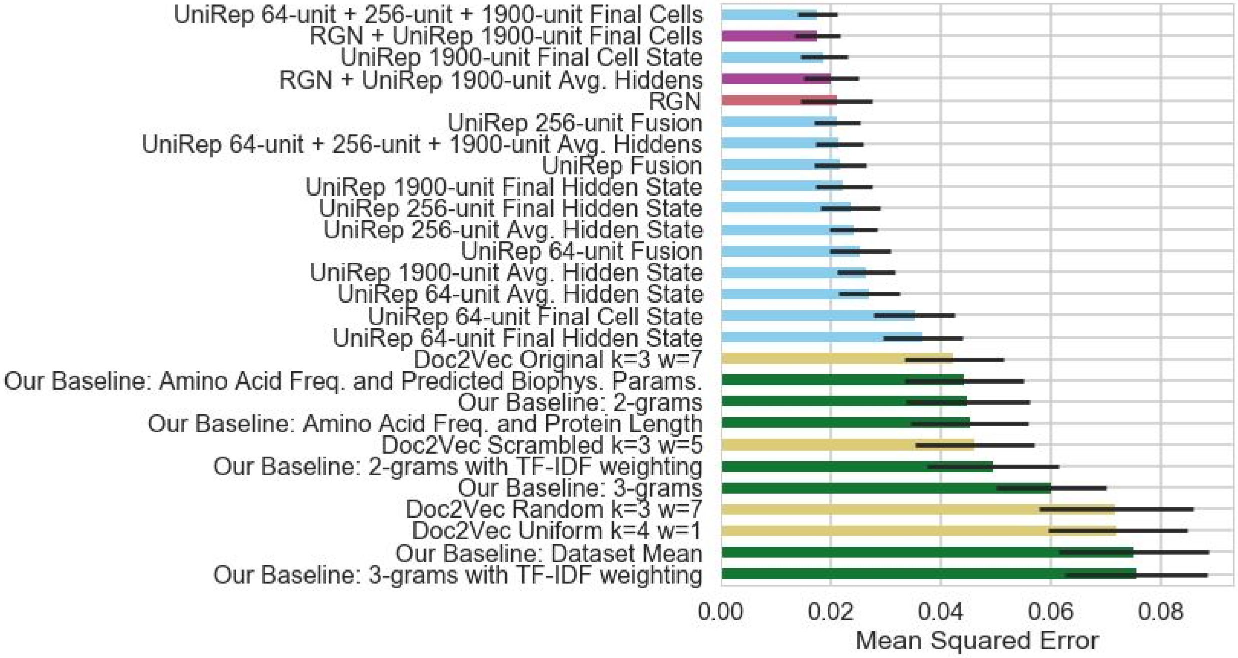
Supplementary Results Figure 13: All representation results in variant effect prediction task task. Subset: Yap65 (WW domain).



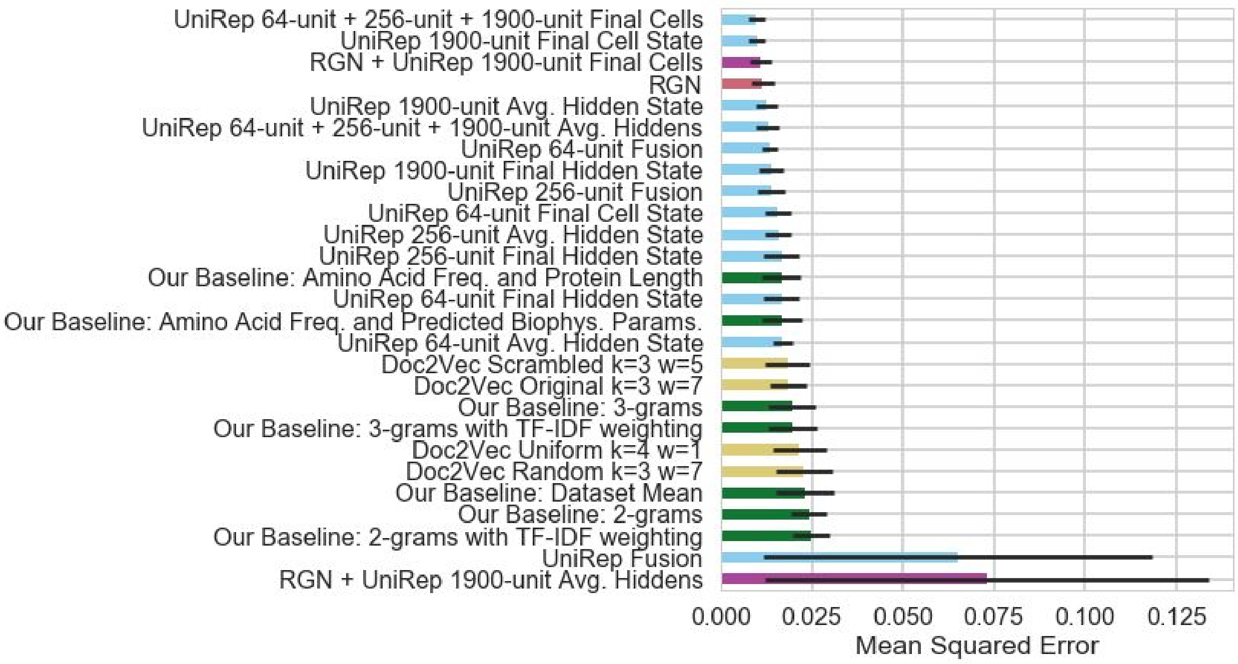
Supplementary Results Figure 14: All representation results in variant effect prediction task task, Combined Data.



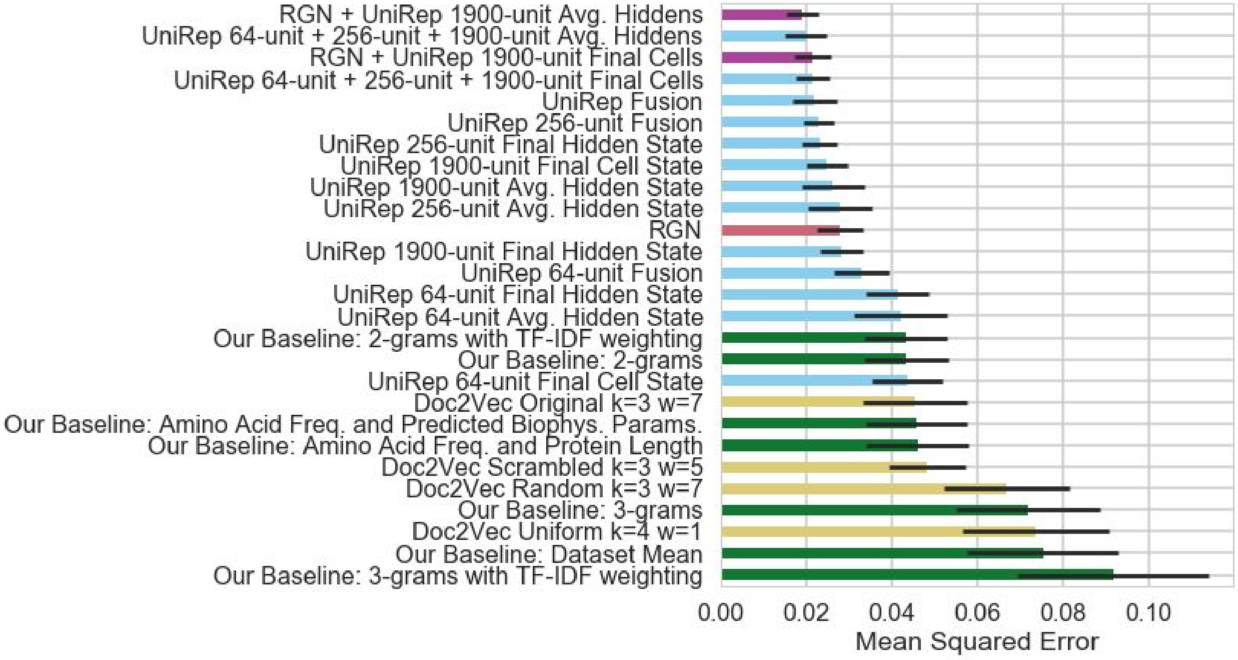
Supplementary Results Figure 15: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EEHEE_rd3_0037.



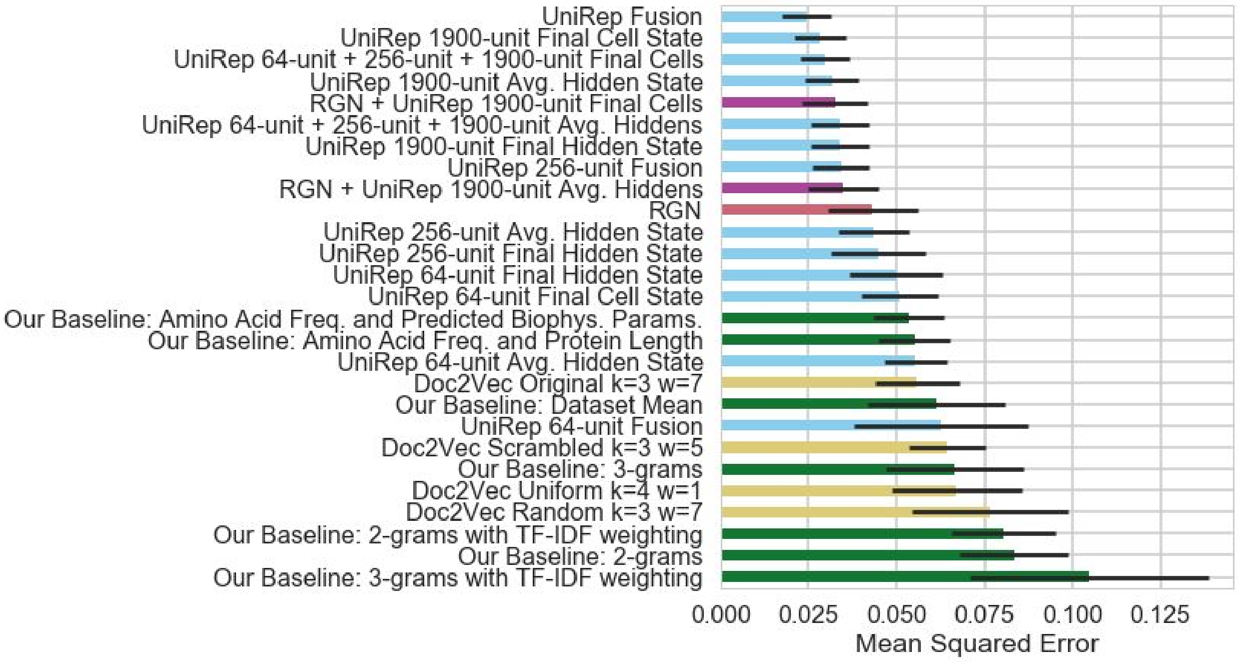
Supplementary Results Figure 16: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EEHEE_rd3_1498.



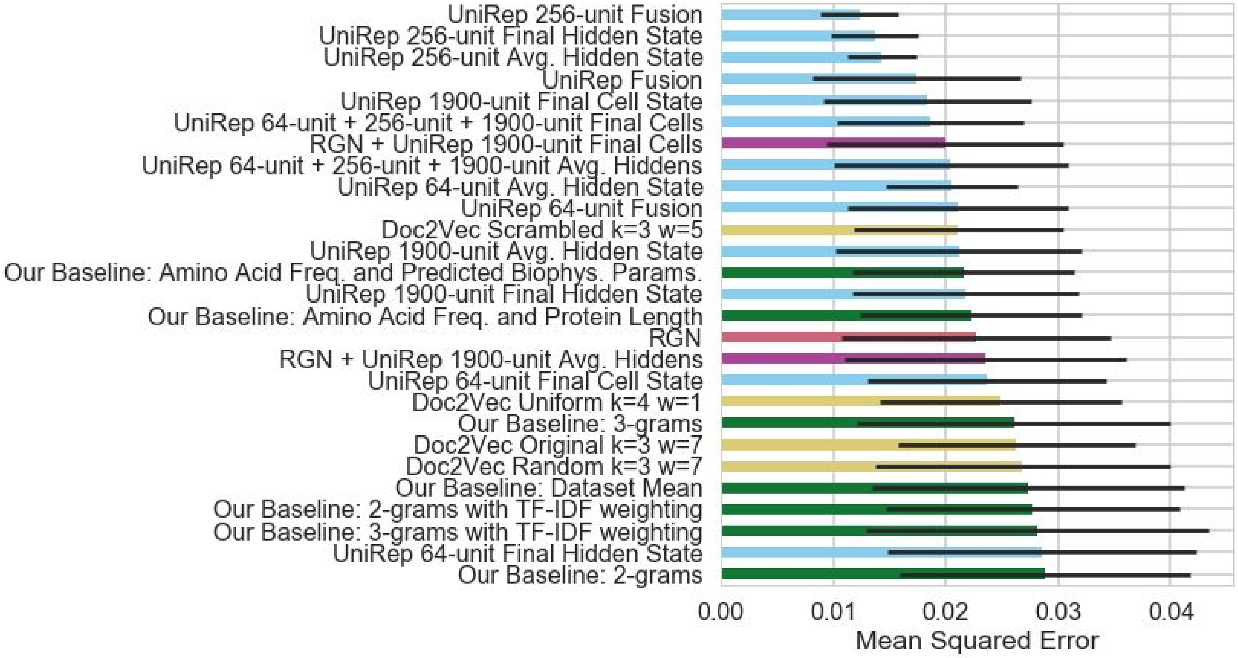
Supplementary Results Figure 17: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EEHEE_rd3_1702.



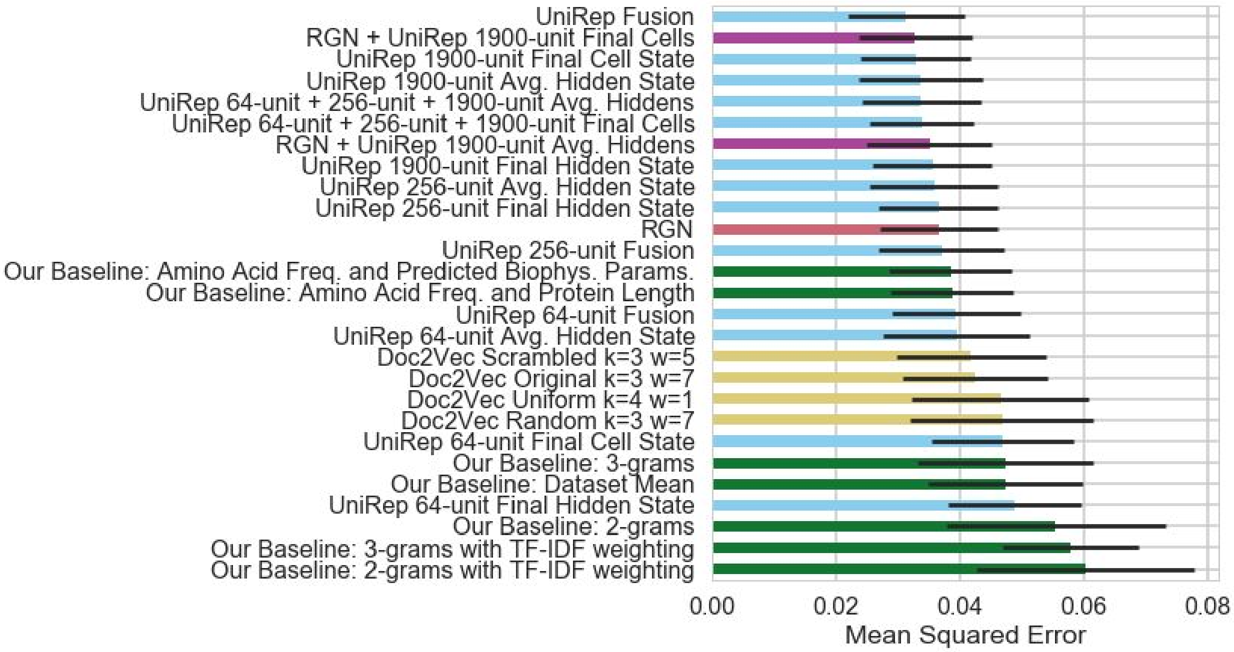
Supplementary Results Figure 18: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EEHEE_rd3_1716.



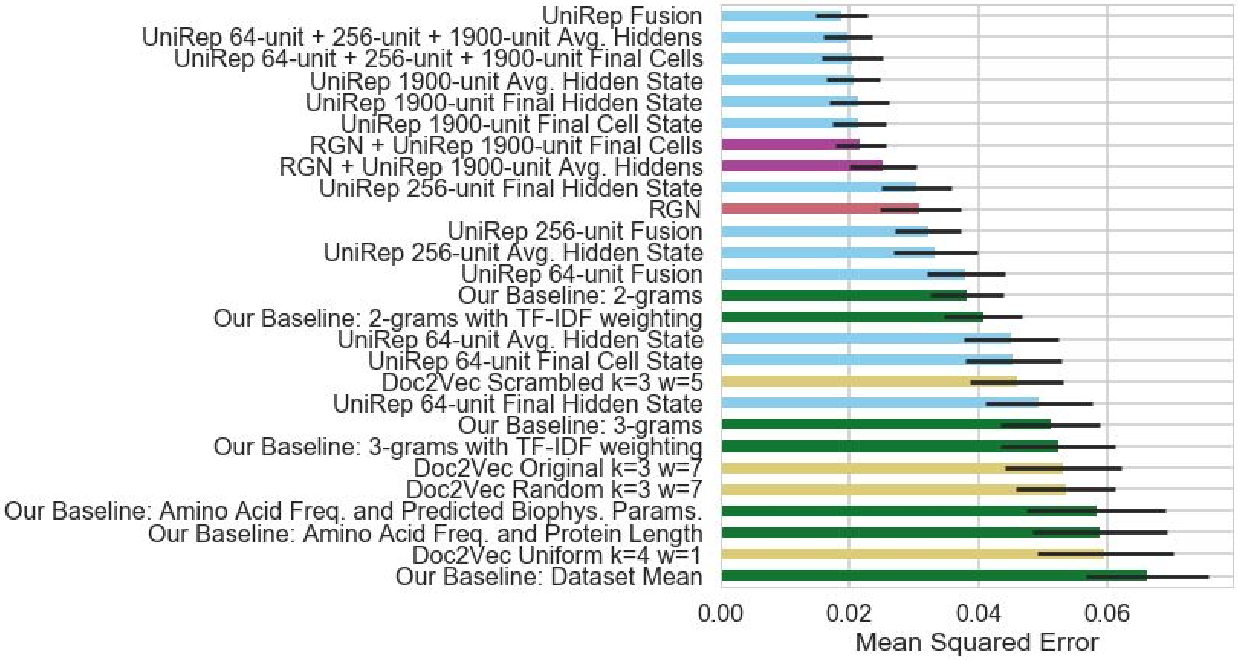
Supplementary Results Figure 19: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EHEE_0882.



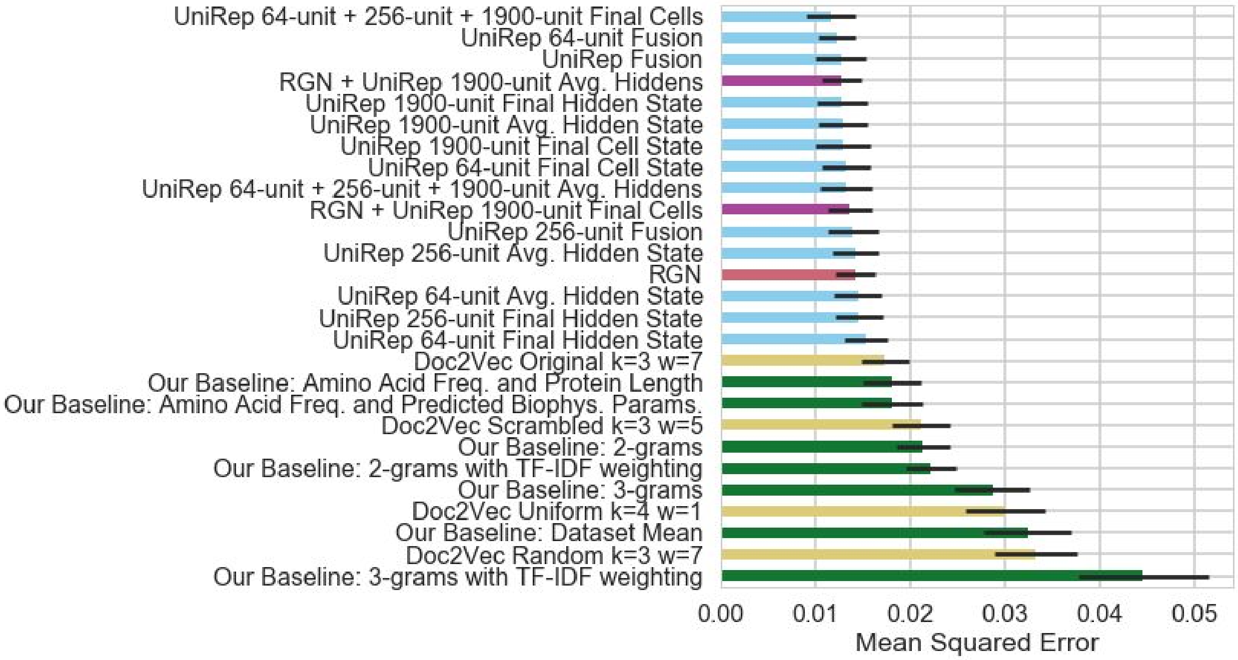
Supplementary Results Figure 20: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EHEE_rd2_0005.



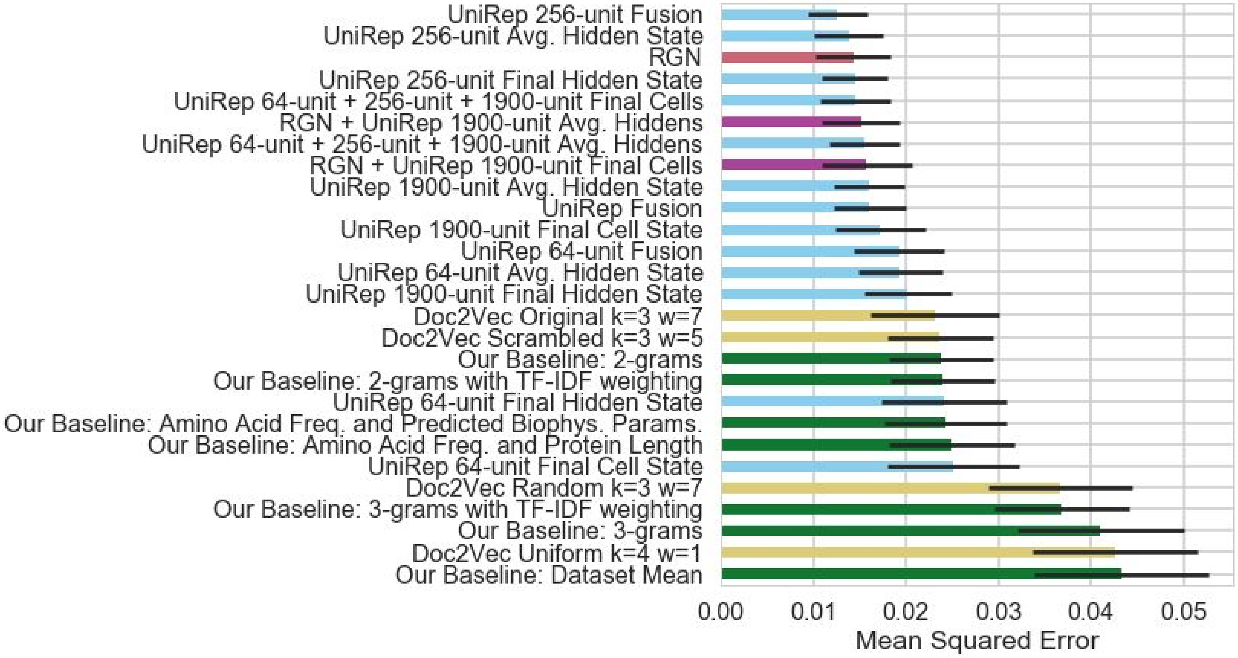
Supplementary Results Figure 21: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: EHEE_rd3_0015.



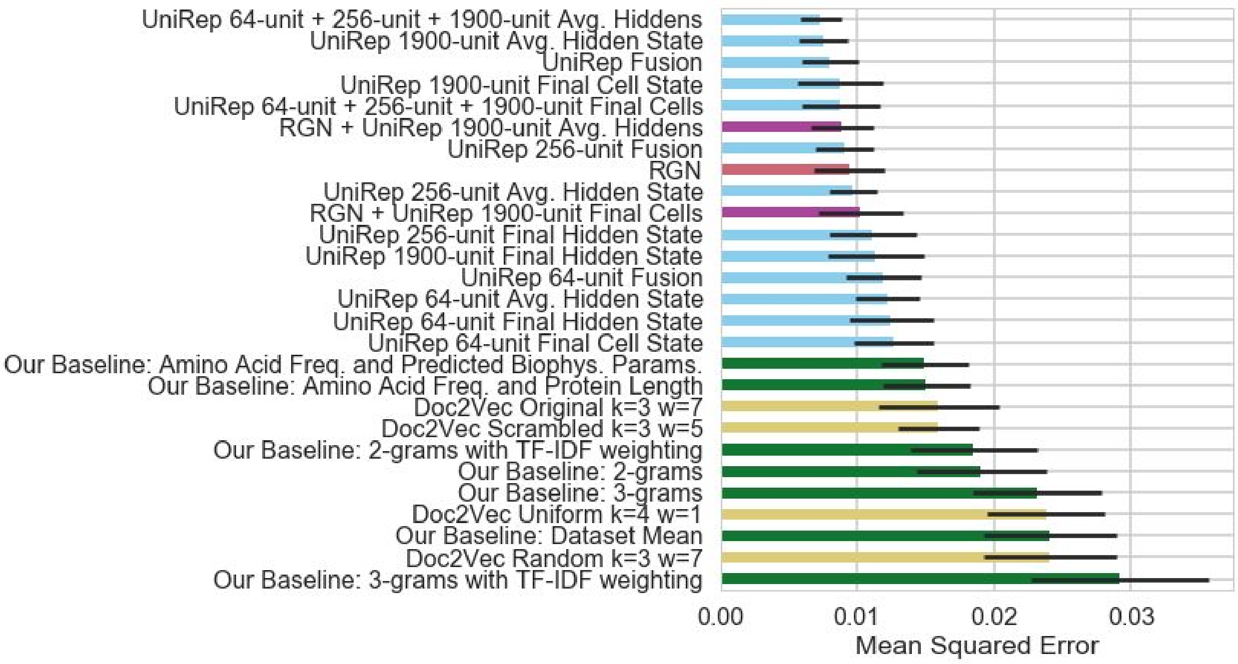
Supplementary Results Figure 22: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HEEH_rd2_0779.



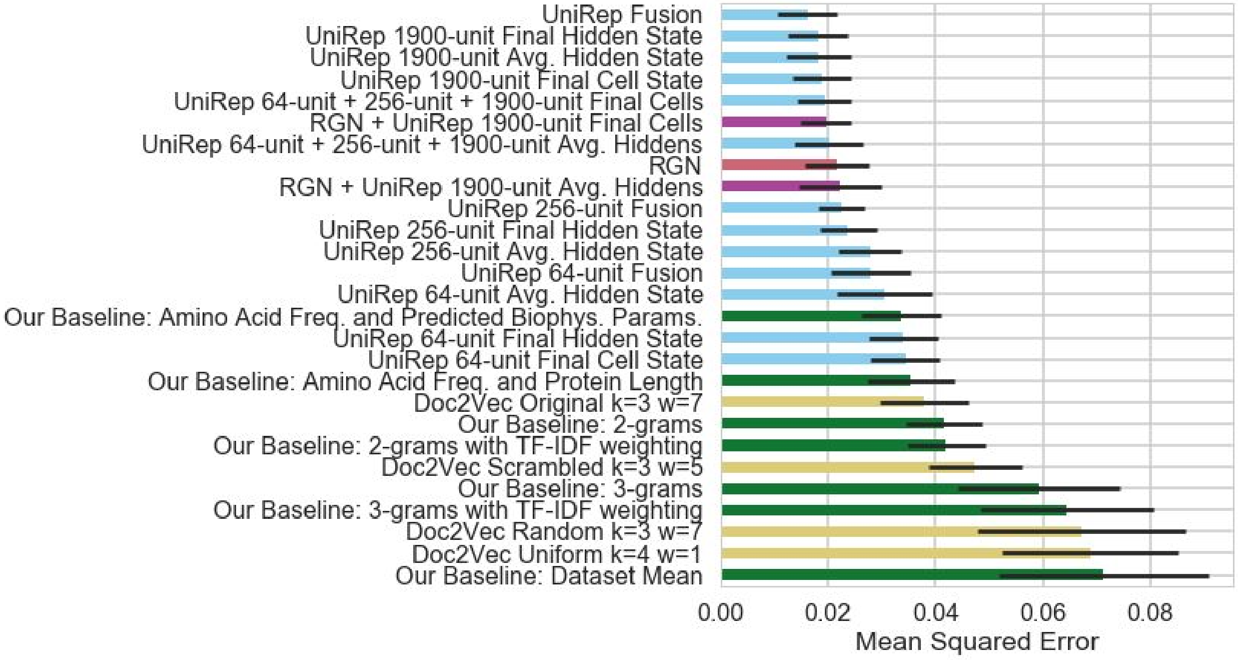
Supplementary Results Figure 23: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HEEH_rd3_0223.



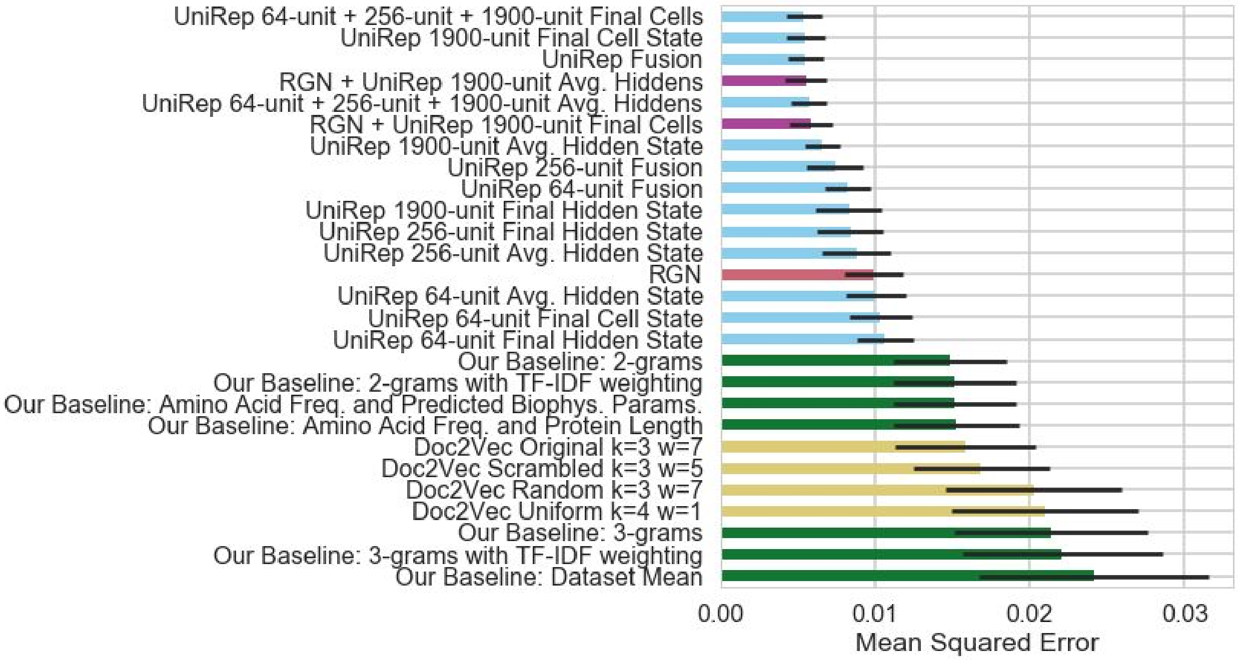
Supplementary Results Figure 24: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HEEH_rd3_0726.



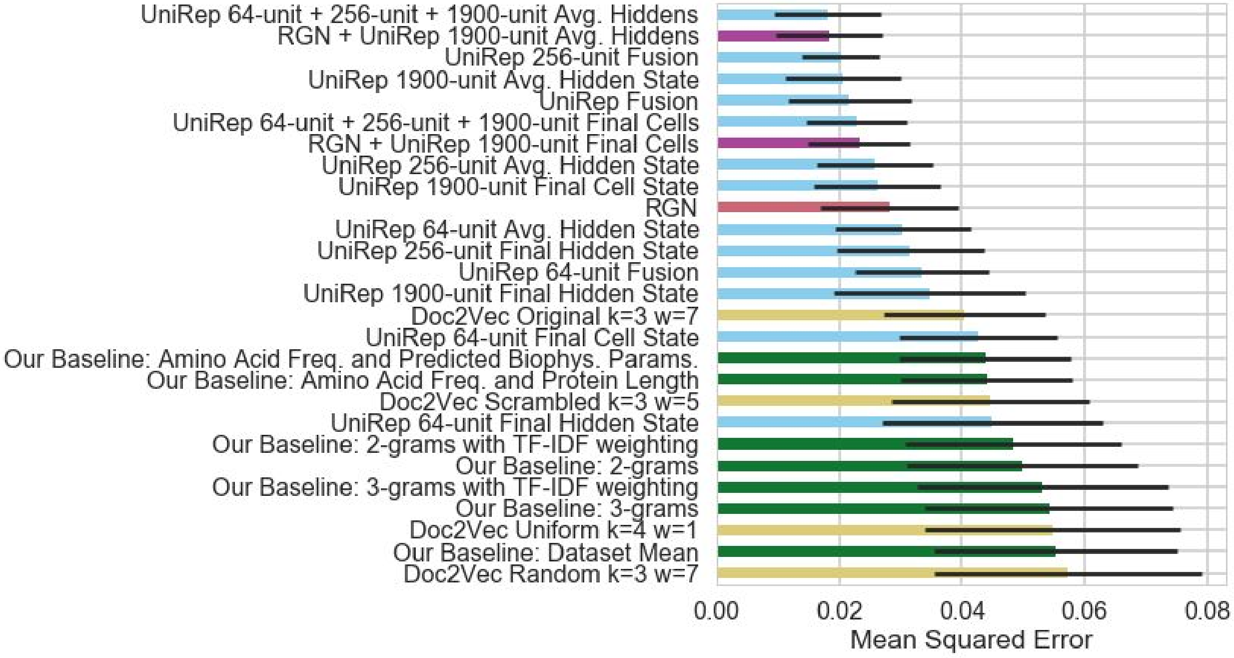
Supplementary Results Figure 25: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HEEH_rd3_0872.



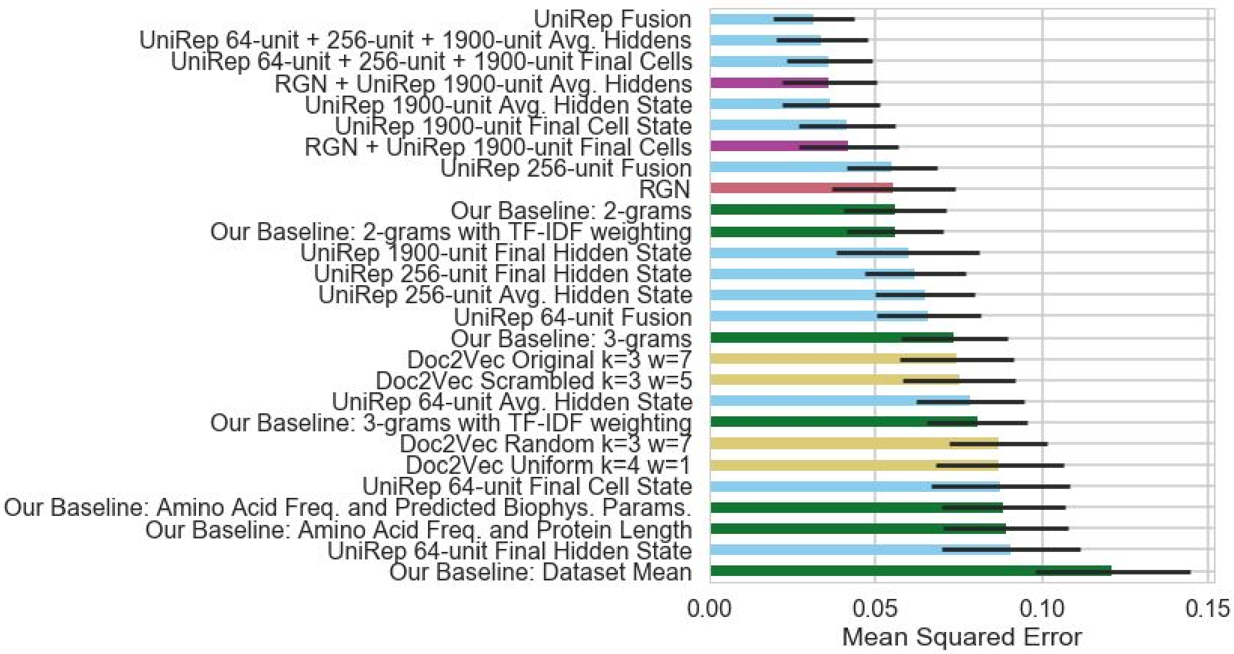
Supplementary Results Figure 26: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HHH_0142.



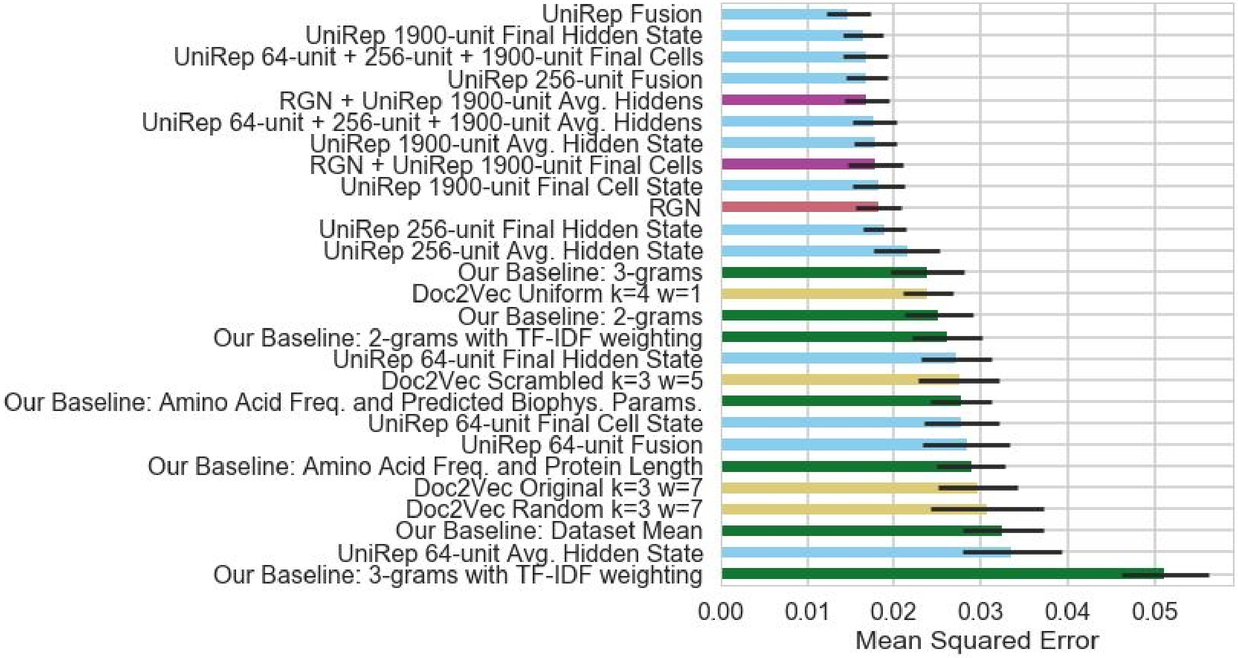
Supplementary Results Figure 27: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HHH_rd2_0134.



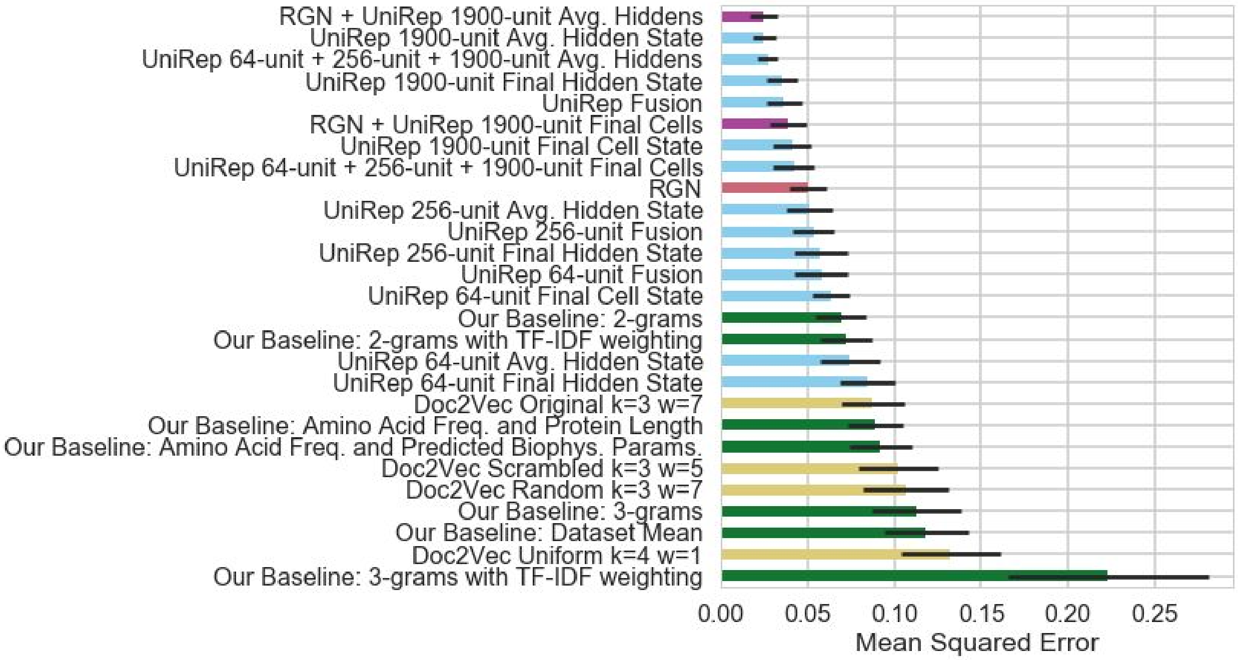
Supplementary Results Figure 28: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: HHH_rd3_0138.



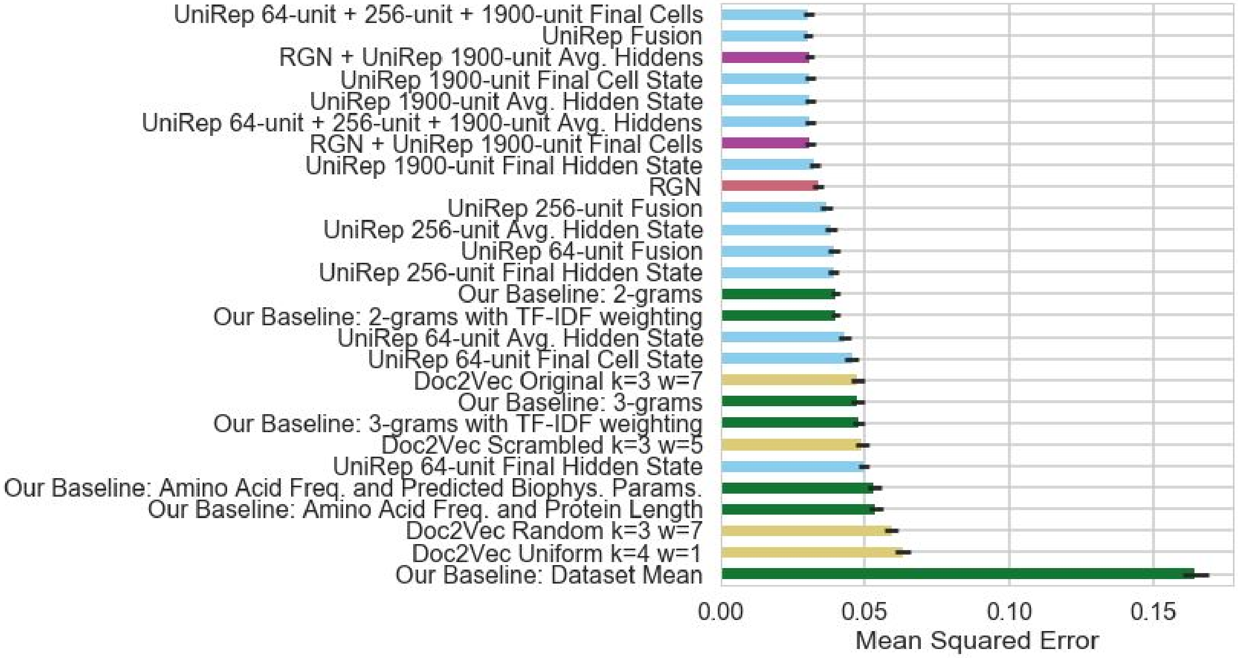
Supplementary Results Figure 29: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: Pin1.



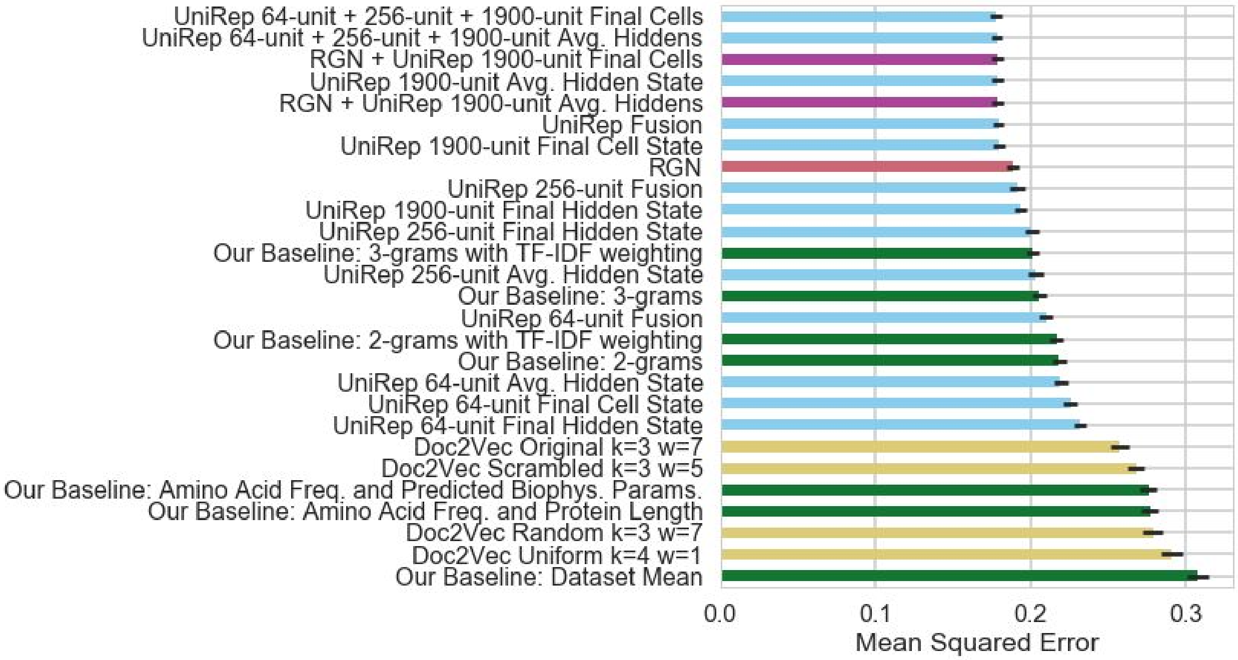
Supplementary Results Figure 30: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: hYAP65.



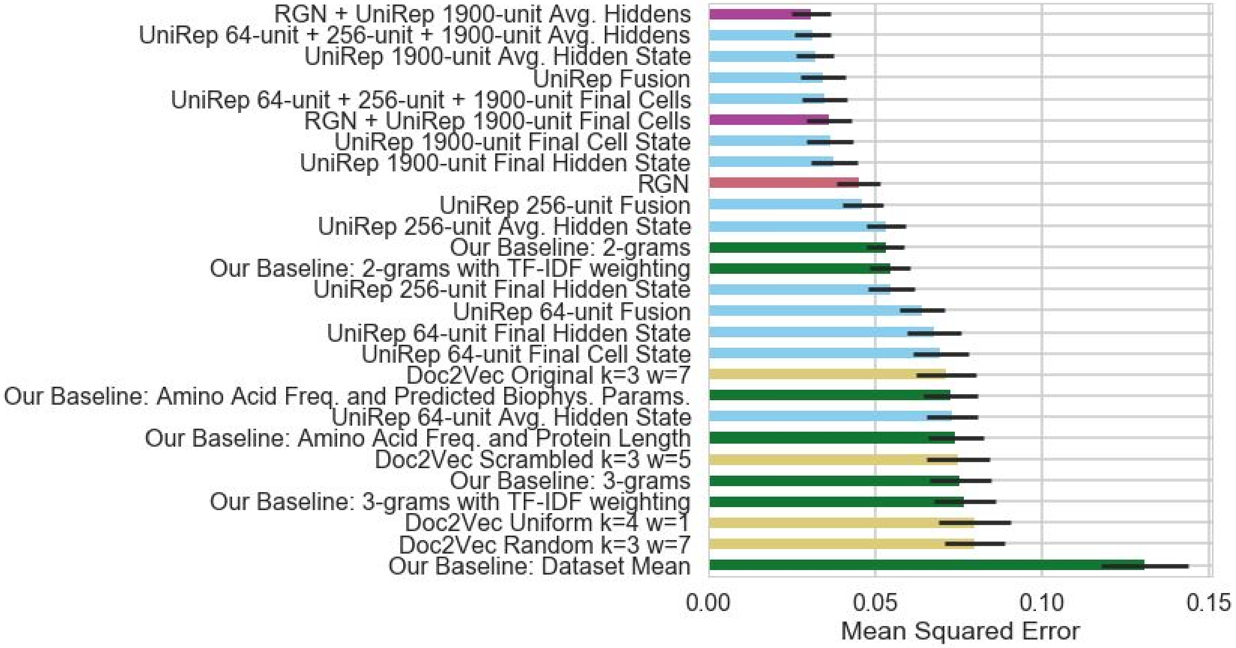
Supplementary Results Figure 31: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task. Subset: villin.



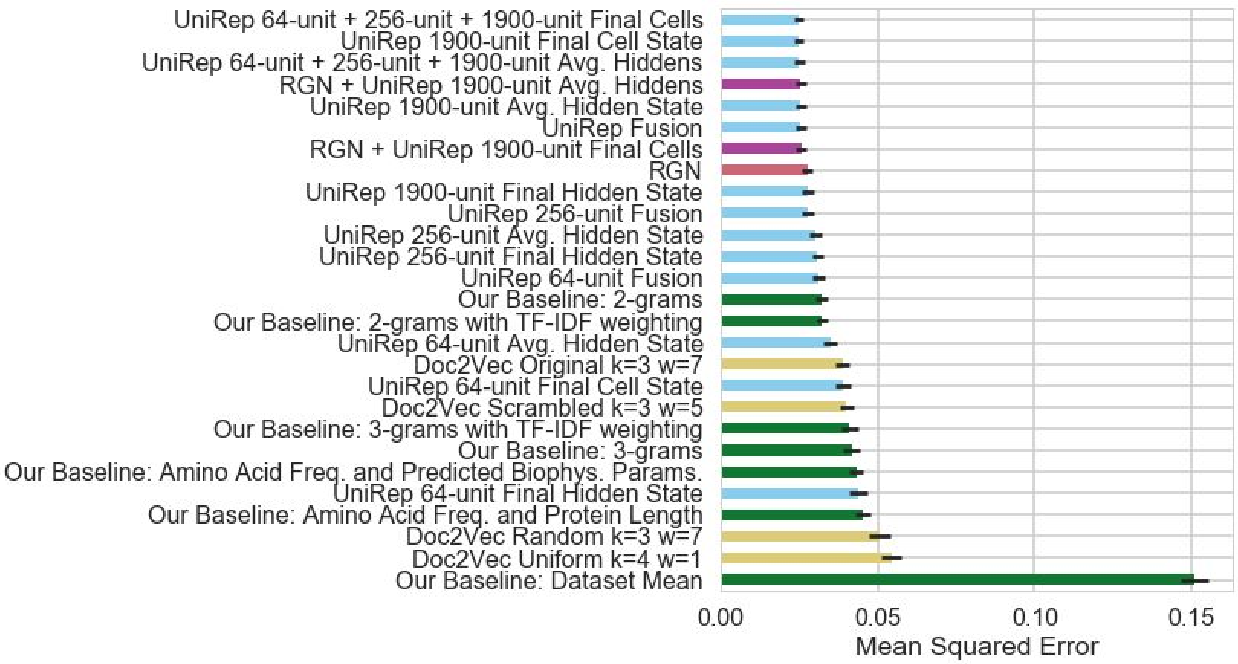
Supplementary Results Figure 32: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis task, Combined Data.



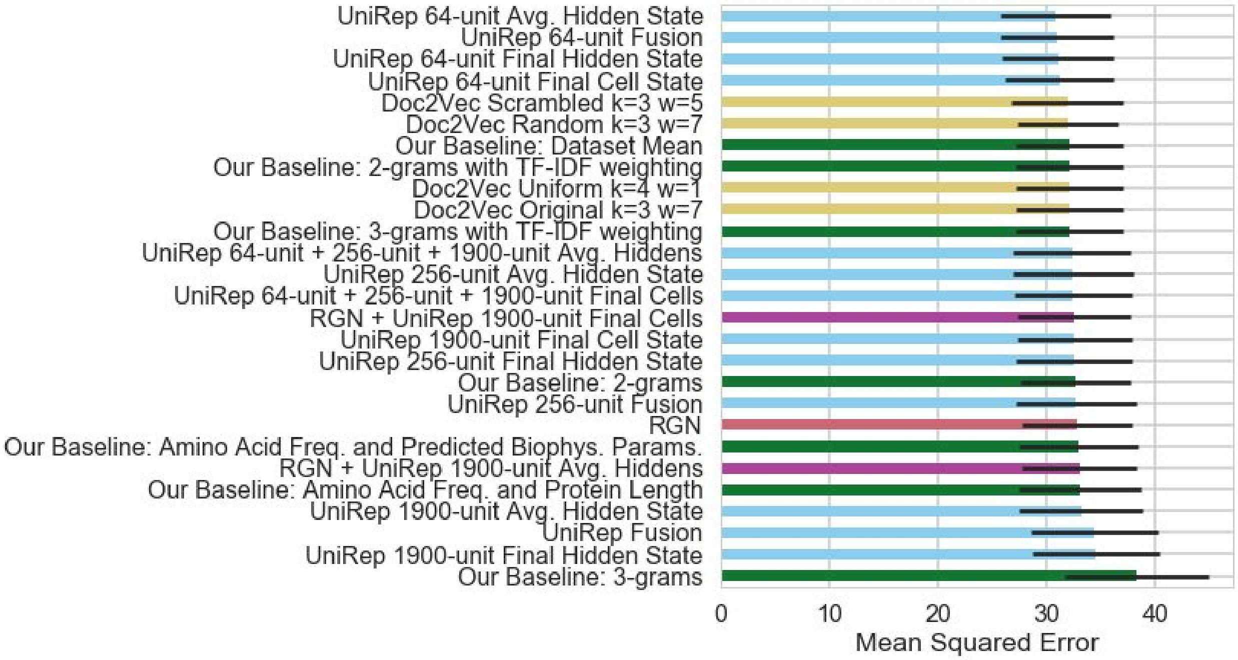
Supplementary Results Figure 33: All representation results in de-novo designed proteins stability - design rounds task.



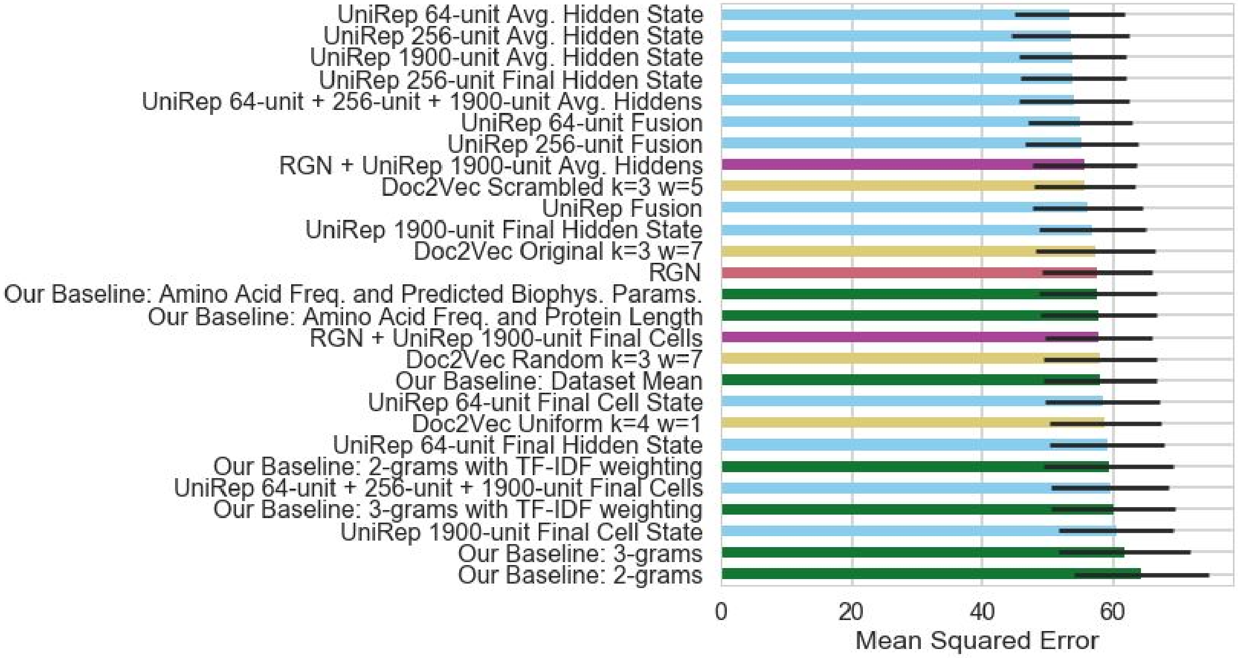
Supplementary Results Figure 34: All representation results in natural & de-novo designed mutant proteins stability - site saturation mutagenesis task. Subset: Natural.



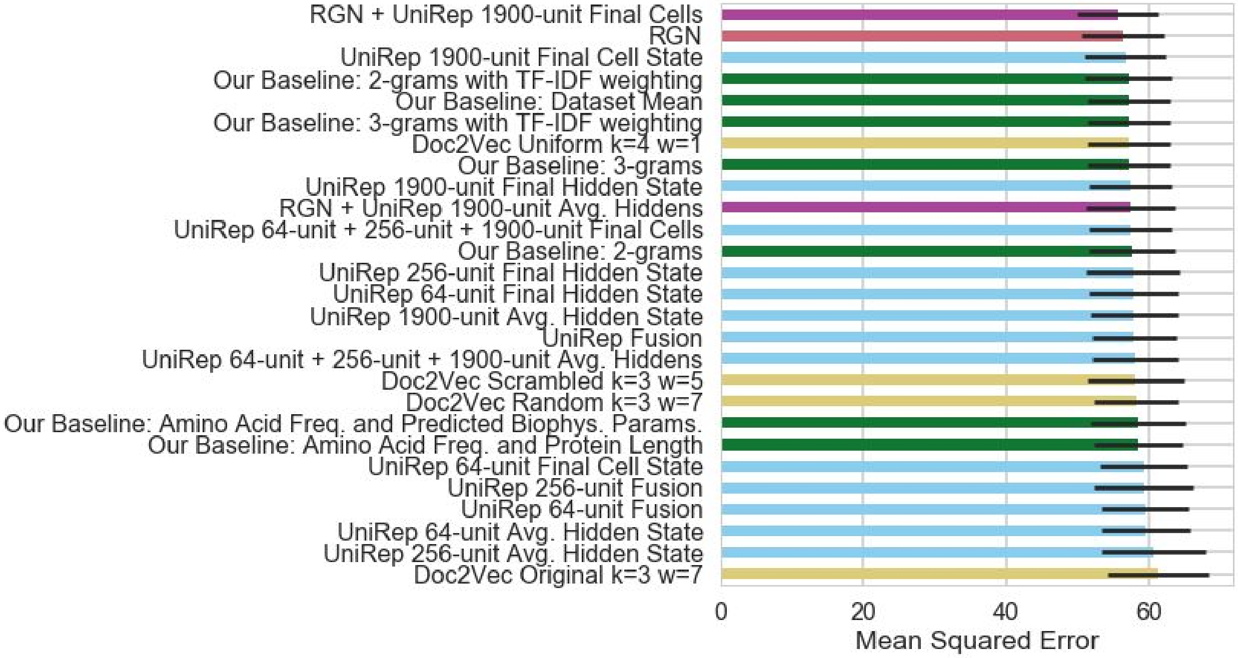
Supplementary Results Figure 35: All representation results in natural & de-novo designed mutant proteins stability - site saturation mutagenesis task. Subset: De-Novo Designed.



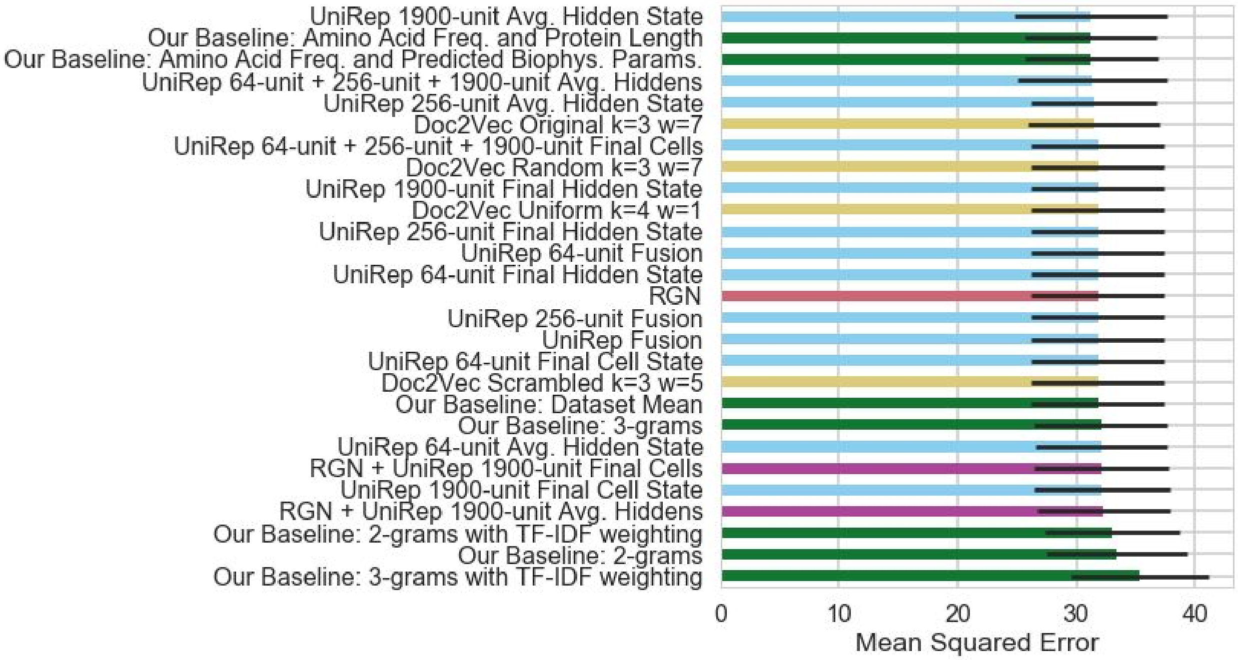
Supplementary Results Figure 36: All representation results in thermostability prediction across organisms task. Subset: H. Sapiens.



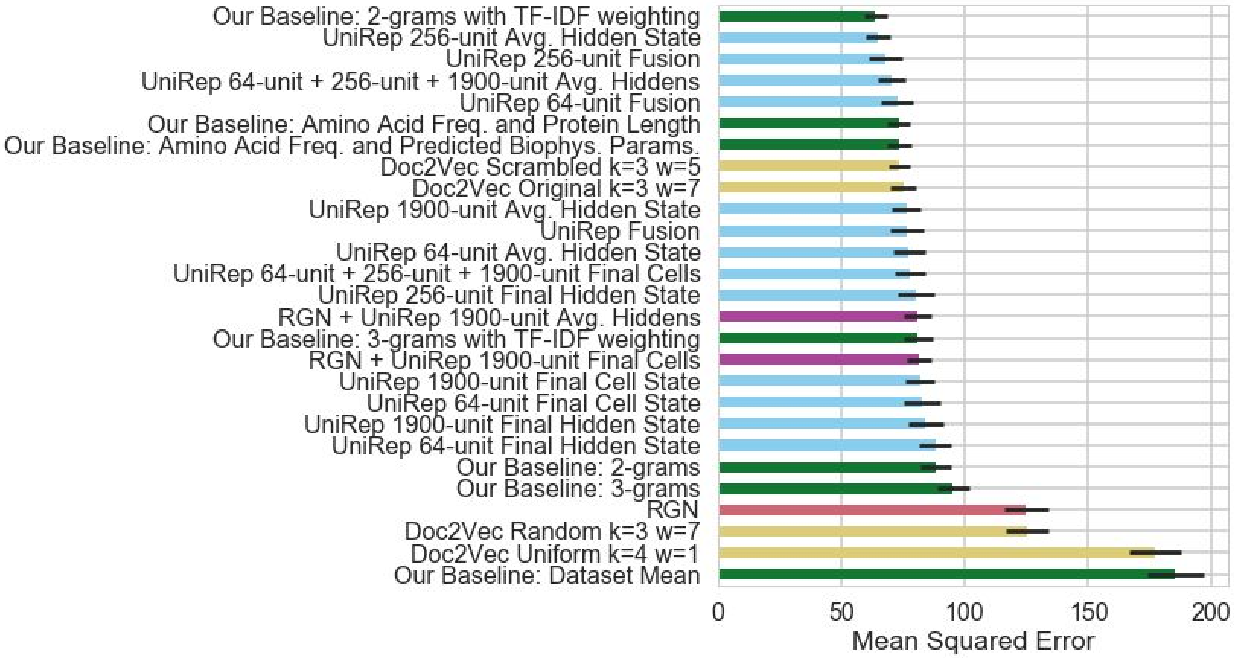
Supplementary Results Figure 37: All representation results in thermostability prediction across organisms task. Subset: E. Coli.



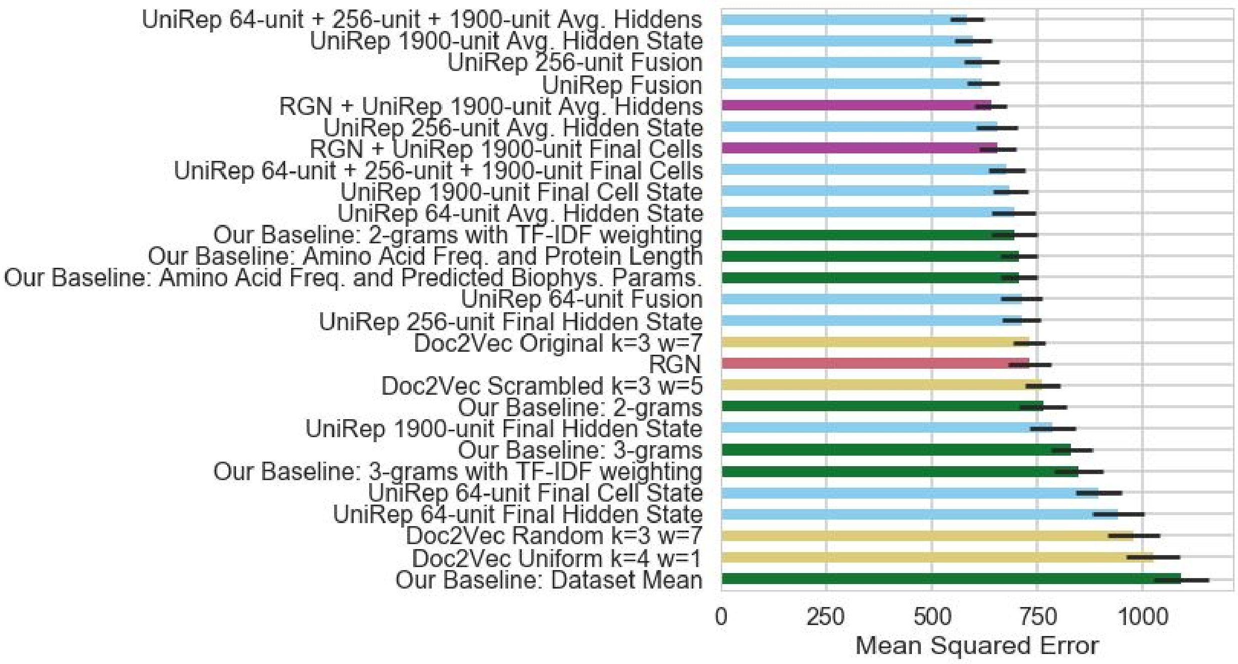
Supplementary Results Figure 38: All representation results in thermostability prediction across organisms task. Subset: S. Thermophilus.



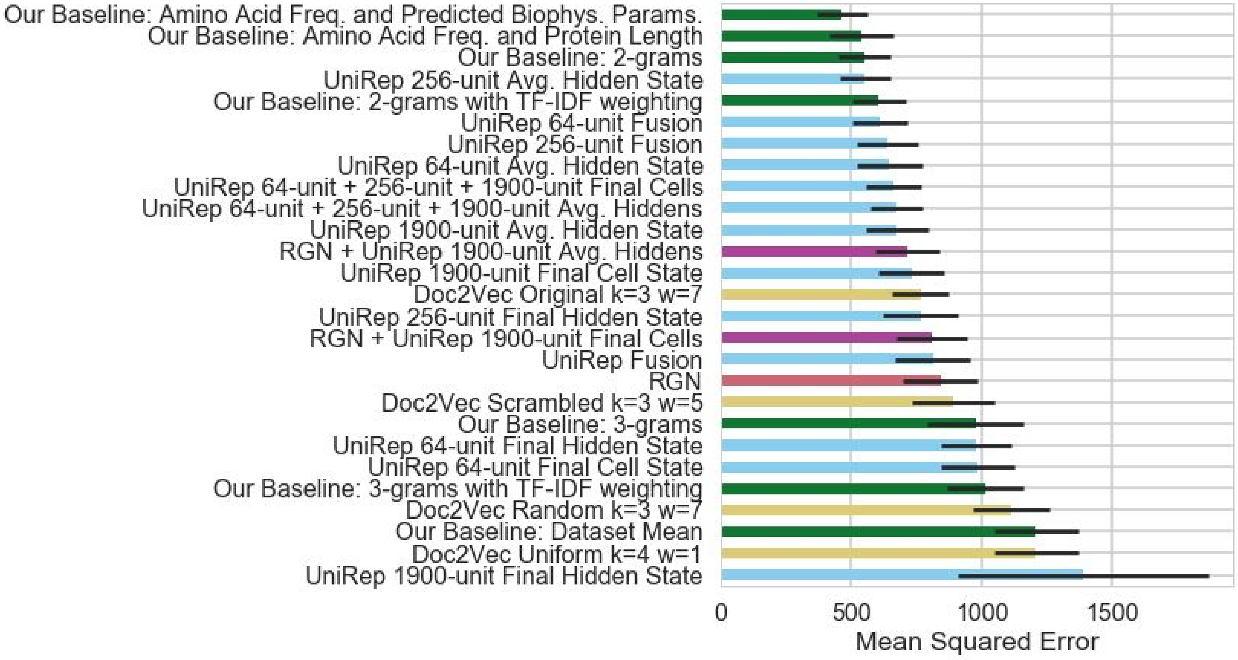
Supplementary Results Figure 39: All representation results in thermostability prediction across organisms task. Subset: S. Cerevisiae.



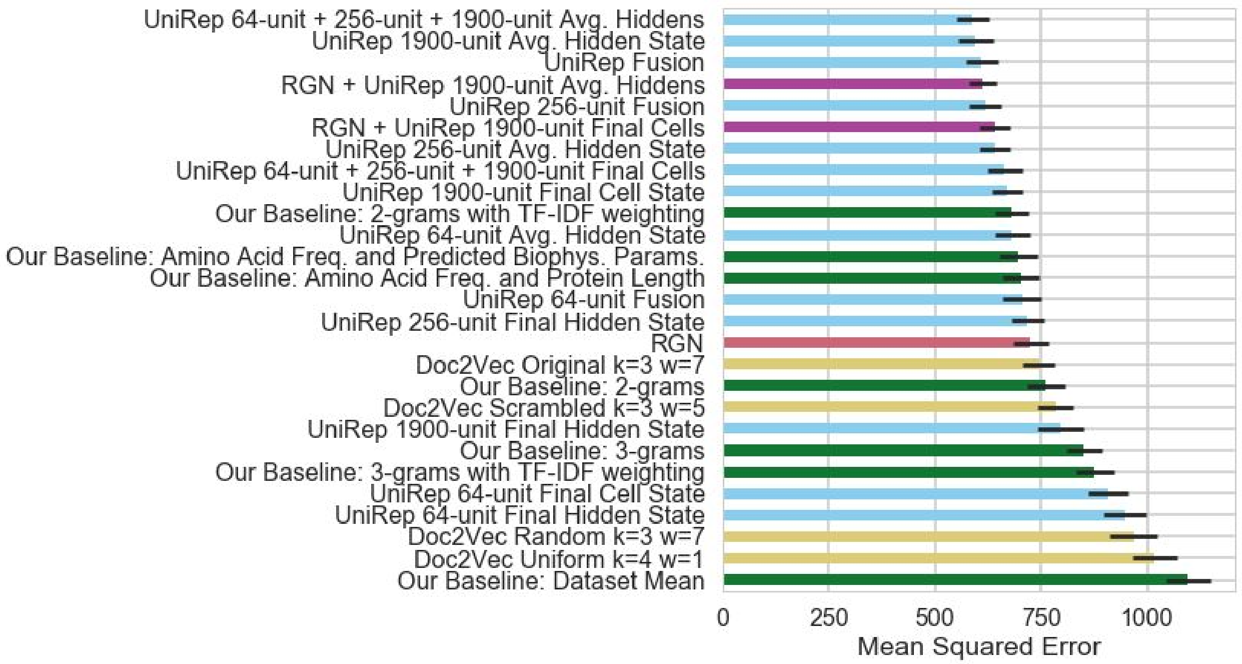
Supplementary Results Figure 40: All representation results in thermostability prediction across organisms task, Combined Data.



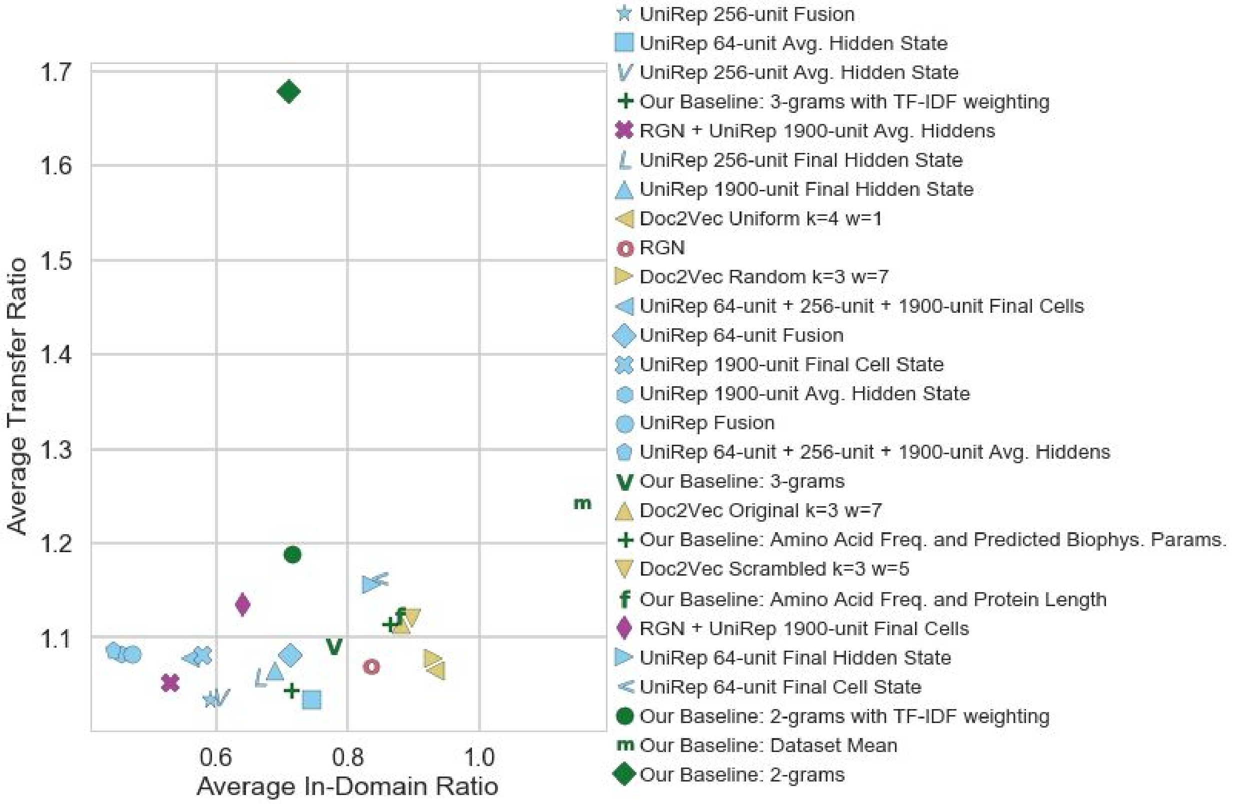
Supplementary Results Figure 41: All representation results in protein solubility prediction task. Subset: E. Coli.



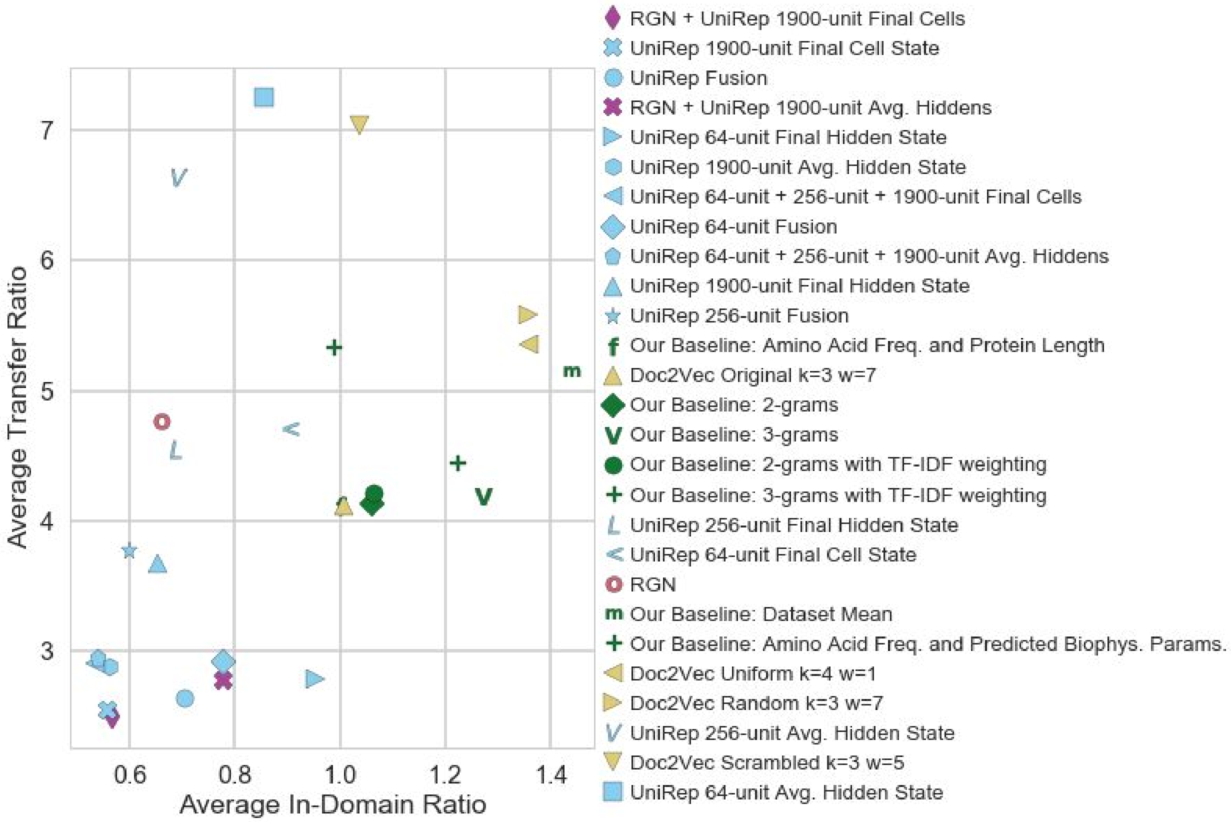
Supplementary Results Figure 42: All representation results in protein solubility prediction task. Subset: S. Cerevisiae.



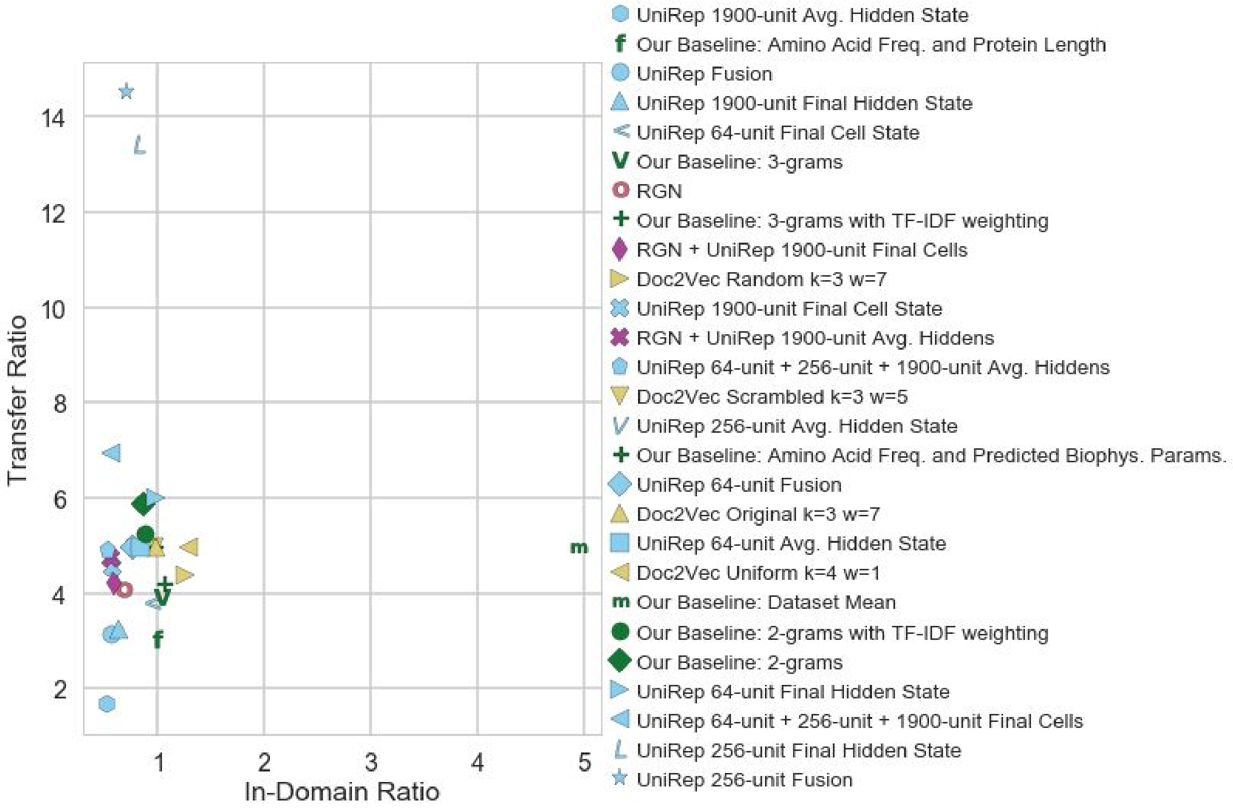
Supplementary Results Figure 43: All representation results in protein solubility prediction task, Combined Data.



Supplementary Figure 44: All representation results in variant effect prediction task transfer task.



Supplementary Figure 45: All representation results in natural & de-novo designed proteins stability - site saturation mutagenesis transfer task.



Supplementary Figure 46: All representation results in natural & de-novo designed mutant proteins stability - remote transfer task.

Inconsistent datasets excluded from benchmarking during preliminary analysis

We considered additional two datasets for our benchmark suite but ultimately decided not to use them. The smaller-scale thermostability prediction dataset from (Leuenberger 2017) with 4 organisms (Thermophilus, E. Coli, Human, S. cerevisiae) and 3275 proteins was discontinued due to the fact that none of representations/baselines that we tried could achieve performance significantly different from that achieved with predicting the dataset mean on any of the organism subsets (Supp. Results Fig. 36-40), indicating low information content, likely resulting from experimental variability. The solubility prediction dataset combines two large-scale solubility-measurement projects (E. coli and S. cerevisiae, 3549 proteins). There we see extreme differences in performance of baselines on these datasets (Supp. Results Fig. 41-43), likely also caused by experimental variability in these projects, which precludes their use as an accurate gauge of representation performance. The variability likely resulted from different tacit experimental decisions in the two projects - while ostensibly using the same methodology in the same lab, they occurred 9 years apart and were performed by different groups of lab members.