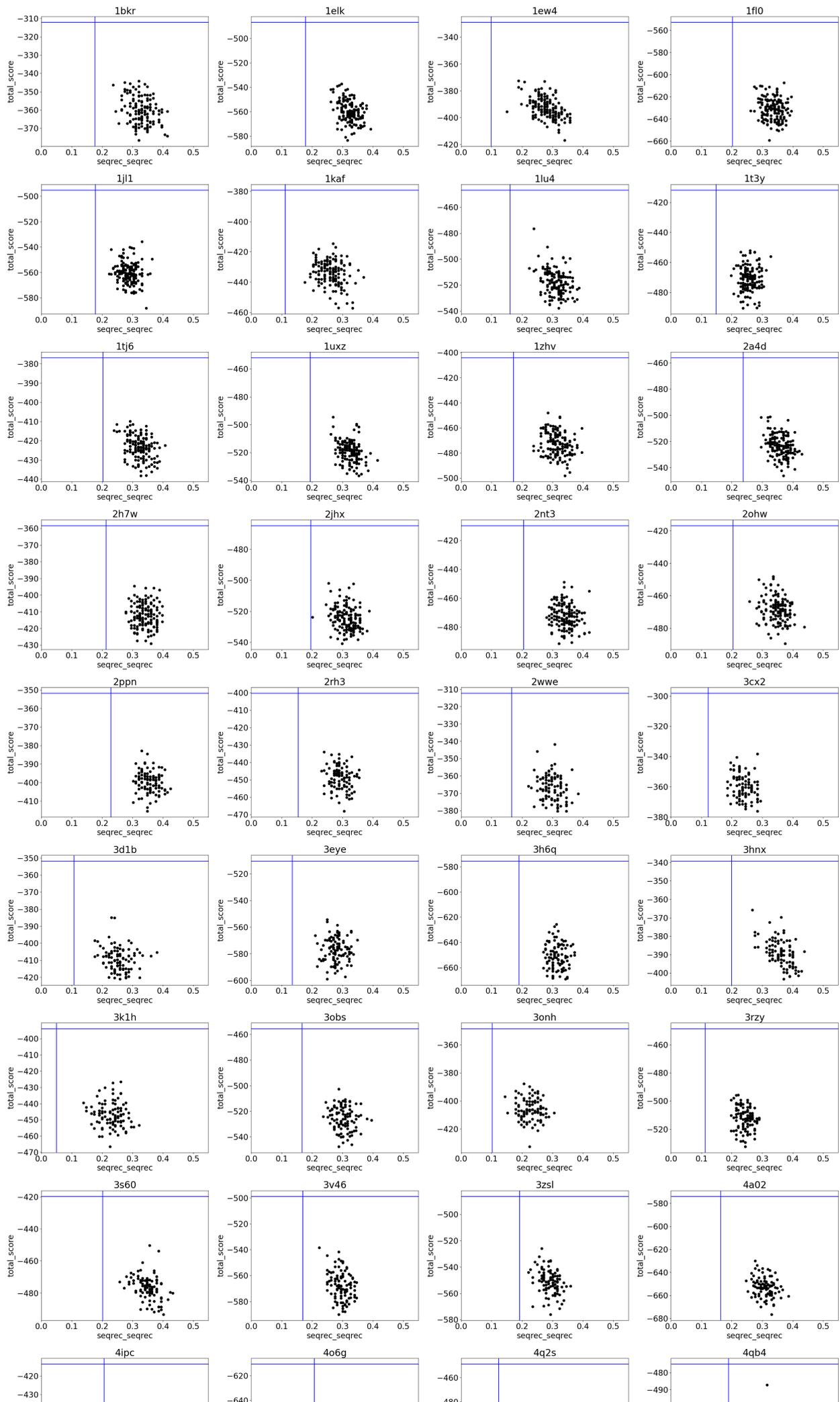# Scientific test: design_fast

FAILURES

None

RESULTS

AUTHOR AND DATE

This benchmark was set up by Julia Koehler Leman (julia.koehler.leman@gmail.com), PI Richard Bonneau, March 2020

Input data and command lines are from Chris Bahl and Jack Maguire.

PURPOSE OF THE TEST

The benchmark tests how well FastDesign can recover native sequences on the benchmark set.

BENCHMARK DATASET

The benchmark set contains 48 proteins between 102 and 176 residues, originally used by Frank DiMaio for his improvements to the energy function. The set covers alpha-helical bundles, beta-sheets proteins and mixed alpha/beta folds.

PROTOCOL

The protocol runs FastDesign in RosettaScripts currently with 1 iteration, nstruct 100, no extrachi. Probably should try 5 iterations as originally suggested. 1 iteration generates a decoy in about 2000 seconds. This makes this protocol run for about 48 x 100 x 2000 / 3600 = 2666 CPU hours.

PERFORMANCE METRICS

We use sequence recovery between the native and the design computed via SimpleMetrics in RosettaScripts. The cutoffs were defined for sequence recovery, for each protein take the minimum minus 2 stdev. For the score, per protein take the maximum plus 5 stdev.

KEY RESULTS

The sequence recovery metric has been used for many years to benchmark design applications. Historically, sequence recoveries are somewhere between 30% and 60% at the maximum. It is difficult for the scorefunction to recapitulate native sequences accurately. It might be worth noting that we do not expect 100% sequence recovery even with a "perfect" energy function and "perfect" optimizer, since evolution optimizes proteins for marginal stability (to allow for degradation) and for other things (function, genetic code, amino acid costs/abundances), while we're trying to optimize for high stability (and maximize the stability of the designed state, without knowing what we're doing to the stability of alternative conformations).

DEFINITIONS AND COMMENTS

LIMITATIONS

The benchmark set only consists of small, soluble proteins. It would be good to know how design performs on larger proteins and more complex folds. For the quality metrics, rotamer recovery could be considered as well.