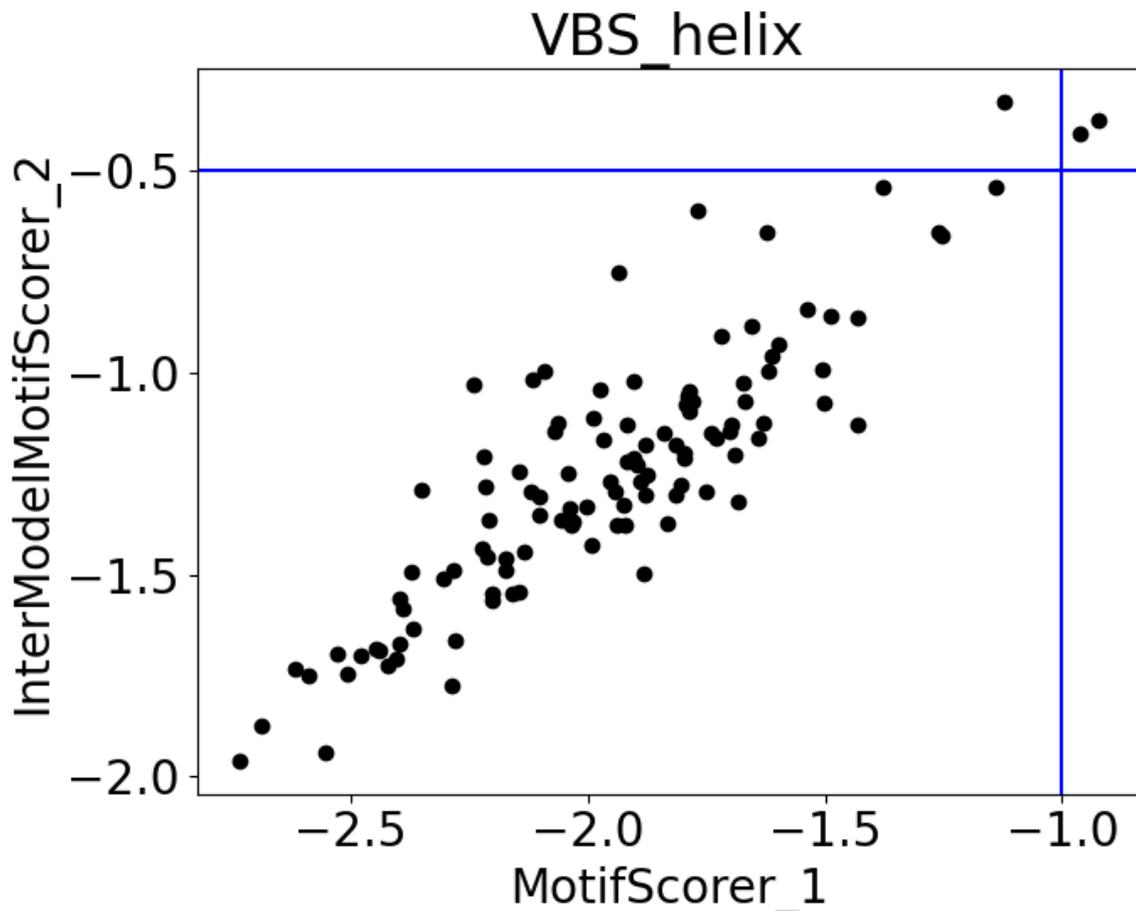


# Scientific test: sewing

## FAILURES

None

## RESULTS



## ## AUTHOR AND DATE

Frank Teets, teetsf@gmail.com, Brian Kuhlman, April 2019

## ## PURPOSE OF THE TEST

This benchmark tests SEWING's capacity to explore restricted conformational spaces and still create long-range interactions; the restricted space magnifies the effect of poor sampling on score, making this the most sensitive test of SEWING's use cases. This tells us whether SEWING is sampling backbone additions with sufficient granularity to fit structure into confined spaces while still placing them in ways that approximate a globular protein.

## ## BENCHMARK DATASET

There are two proteins in this set, vinculin (pdb 1T01, chain A) and the VBS1 helix of talin. (pdb 1T01, chain B) However, only one (VBS1) is actually used as the starting node for SEWING and consequently scored. The other serves as a SEWING "partner protein"; in normal use, this is something to which the starting node is intended to bind and therefore a region of space into which SEWING should not design structure, as well as serving as a target for encouraging the placement of backbone in regions likely to be conducive to expanding that interface via residue-level design. The requisite PartnerMotifScorer is not included in this test set, however; see PERFORMANCE METRICS below. The only other SEWING-specific input file is the segment file created as described in the paper from the TOP8000 dataset (Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V. , Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S. , Lewis, S. M., Arendall, W. B., Snoeyink, J. , Adams, P. D., Lovell, S. C., Richardson, J. S. and Richardson, D. C. (2018), MolProbity: More and better reference data for improved all-atom structure validation. Protein Science, 27: 293-315. doi: 10.1002/pro.3330)

## ## PROTOCOL

The protocol is the AppendAssemblyMover protocol as described in (Protocols for Requirement-Driven Protein Design in the Rosetta Modeling Program (Guffy et al, 2018)) less the residue-level design step. It should run in 75 CPU hours.

## ## PERFORMANCE METRICS

The performance metrics are the MotifScorer and InterModelMotifScorer. MotifScore measures whether the added structure is designable, and IMMS determines if it's globular. MotifScore represents the sum total of the highest possible full-atom attractive energy for any two small hydrophobic residues placed at every pair of residue positions within 6Å of each other, normalized by the length of the design. This serves as a quickly evaluable metric of how well residue-level design can theoretically perform in designing a core from these residue positions. InterModelMotifScore restricts its evaluation to pairs of residues separated by at least one helix, meaning that it excludes all interactions between residues on the same helical hairpin. This is used internally to bias sampling toward globular proteins rather than flat sheets of helices, the latter maximizing the available volume for adding new structure. It also serves as a metric of globularity, and particularly of whether a given backbone has enough high-contact-order interactions for residue-level design to stabilize it in the designed configuration.

The cutoffs are defined as MotifScorer < -1.0/residue and InterModelMotifScorer < -0.5/residue, evaluated for the top 10% of outputs; the cutoffs are specific to this test case, and come from the paper. MotifScore is always numerically less than InterModelMotifScore for a given design, since InterModelMotifScore runs the same score function on a subset of the residue pairs, but if MotifScore passes the threshold and InterModelMotifScore does

not, that indicates that SEWING has failed to sample with sufficient granularity to produce a globular protein; if MotifScore itself fails to reach  $< -1.0/\text{residue}$ , that generally indicates that SEWING has either added nothing at all to the starting node or that its additions have resulted in an unstructured assembly of helices distant from each other in space. This generally happens in a small fraction of decoys for every design run, but if it happens consistently (enough that  $>90\%$  of the runs fail the test) it usually indicates that SEWING has failed to sample the restricted space around the starting node sufficiently.

The PartnerMotifScorer has been excluded from this test as it is less predictive than the other two, since it implicitly assumes that double-sided design is possible and will therefore position helices irrespective of the amino acids at the interface. Internally it serves to bias SEWING toward backbones suitable for interface design, but as a metric, good PartnerMotifScore is not predictive of good interface in the same way that (InterModel)MotifScore predicts a designable core so the score value is not informative. Similarly, since the output of SEWING precedes residue-level design, it includes the native residues from the helices it adds, and these will almost certainly clash with each other. It is therefore not appropriate to evaluate the full-atom score of the design as produced by SEWING alone, for which reason only internal SEWING scores are included in the benchmark.

## **## KEY RESULTS**

As this is a protocol for design, the baseline is relative to the original SEWING described in (Design of structurally distinct proteins using strategies inspired by evolution. (Jacobs et al.,2016)). It is expected that a small proportion of runs will fail to add anything to the starting helix, producing a population of data points close to 0 in both scores. Outside of that population, MotifScore and InterModelMotifScore should correlate as indicated by the cutoffs, indicating that SEWING is both placing helices designably close to each other (which is measured by MotifScore) and that it is also doing so for helices not adjacent in sequence space (as measured by InterModelMotifScore.)

## **## DEFINITIONS AND COMMENTS**

Run with 1000 minimum and maximum cycles, 9 minimum and maximum segments, a hash window width of 4, and a MotifScore:InterModelMotifScore weight ratio of 1:10, as described in the paper.

## **## LIMITATIONS**

This protocol is intentionally agnostic to idiosyncrasies in the correlation between (InterModel)MotifScore and final full-atom score; a future, multistage-compatible SEWING could be better benchmarked by actually relaxing those structures that passed the post-SEWING filters.