RNAnneal-ss scaffold-seeded sampling: representative50 (under300)

Scope

This report compares **RNAnneal-ss** against **RNAstructure**, **LinearFold-V**, and **EternaFold** on a **50-target** subset of the FR3D/BGSU under400 benchmark (truth length ≤ 300), stratified into **50-nt length buckets**.

Methods (what differs)

- **Baselines:** RNAstructure (`AllSub`), LinearFold-V (`--zuker`), EternaFold (MEA + `contrafold sample`).

- **RNAnneal-ss variants:** identical CaCoFold + refinement + MCMC pipeline, but the single-strand scaffold generator (`Fold` + `AllSub`-like suboptimals) is swapped:

  - **RNss(RS):** RNAstructure scaffolds.
  - **RNss(LF):** LinearFold-V scaffolds.
  - **RNss(EF):** EternaFold scaffolds.

- **Duplex sampling:** still uses RNAstructure `DuplexFold` in all RNAnneal-ss variants (LF/EF do not provide duplex sampling).

- **Seeding:** `AllSub`-like suboptimal structures from the chosen scaffold backend are *injected into the RNAnneal-ss refined scaffold set before MCMC*, so they act as actual MCMC starting scaffolds (not a post-hoc merge).

Results (@1 and @100)

Table 1: Overall performance on representative50 (N=50; metrics @1 and best-of-100).

Method	Mean F1@1	Med F1@1	Mean F1@100	Med F1@100	ΔF1	Fail@100
RNss(RS)	0.019	0.000	0.031	0.000	0.013	96.0%
RNss(LF)	0.019	0.000	0.019	0.000	0.000	98.0%
RNss(EF)	0.147	0.000	0.180	0.000	0.034	80.0%
RNAstr	0.516	0.460	0.563	0.506	0.047	4.0%
LF-V	0.519	0.498	0.684	0.701	0.165	2.0%
EFold	0.549	0.546	0.719	0.734	0.170	0.0%

Table 2: Performance by truth-length bucket (50-nt buckets from the manifest).

| Bucket	Method	N	Mean F1@1	Mean F1@100	Fail@100
30-79	RNss(RS)	12	0.078	0.131	83.3%
30-79	RNss(LF)	12	0.080	0.080	91.7%
30-79	RNss(EF)	12	0.220	0.299	66.7%
30-79	RNAstr	12	0.641	0.801	8.3%
30-79	LF-V	12	0.633	0.767	8.3%
30-79	EFold	12	0.621	0.823	0.0%
80-129	RNss(RS)	10	0.000	0.000	100.0%
80-129	RNss(LF)	10	0.000	0.000	100.0%
80-129	RNss(EF)	10	0.471	0.543	40.0%
80-129	RNAstr	10	0.700	0.744	0.0%
80-129	LF-V	10	0.738	0.845	0.0%
80-129	EFold	10	0.668	0.872	0.0%
130-179	RNss(RS)	10	0.000	0.000	100.0%
130-179	RNss(LF)	10	0.000	0.000	100.0%
130-179	RNss(EF)	10	0.000	0.000	100.0%
130-179	RNAstr	10	0.457	0.457	10.0%
130-179	LF-V	10	0.470	0.658	0.0%
130-179	EFold	10	0.529	0.677	0.0%
180-229	RNss(RS)	9	0.000	0.000	100.0%
180-229	RNss(LF)	9	0.000	0.000	100.0%
180-229	RNss(EF)	9	0.000	0.000	100.0%
180-229	RNAstr	9	0.333	0.333	0.0%
180-229	LF-V	9	0.313	0.528	0.0%
180-229	EFold	9	0.396	0.548	0.0%
230-279	RNss(RS)	7	0.000	0.000	100.0%
230-279	RNss(LF)	7	0.000	0.000	100.0%
230-279	RNss(EF)	7	0.000	0.000	100.0%
230-279	RNAstr	7	0.383	0.383	0.0%
230-279	LF-V	7	0.337	0.562	0.0%
230-279	EFold	7	0.497	0.611	0.0%
280-300	RNss(RS)	2	0.000	0.000	100.0%
280-300	RNss(LF)	2	0.000	0.000	100.0%
280-300	RNss(EF)	2	0.000	0.000	100.0%
280-300	RNAstr	2	0.420	0.420	0.0%
280-300	LF-V	2	0.550	0.643	0.0%
280-300	EFold	2	0.499	0.683	0.0%

1 - CDFs
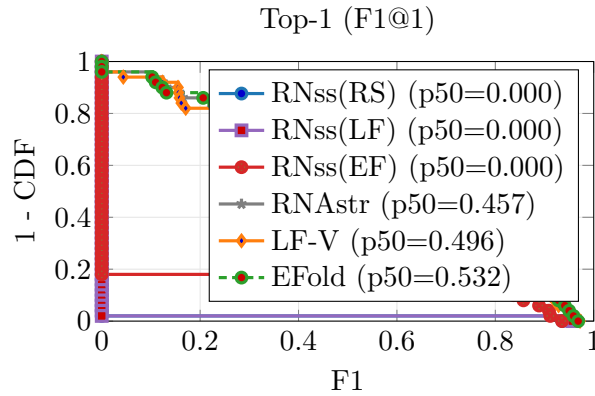
Figure 1: *
F1@1
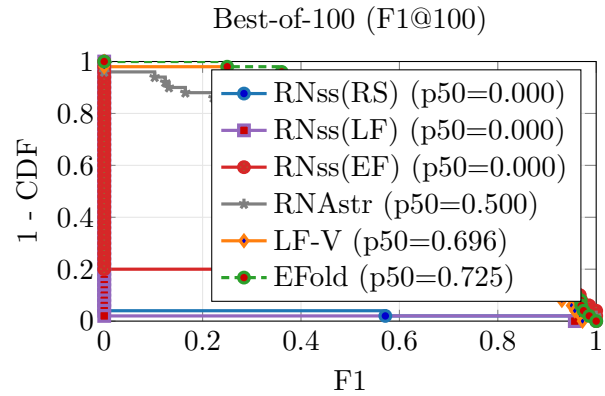
Figure 2: *
F1@100

Figure 3: Empirical survival curves (1 - CDF). Legend includes the median (p50) F1.
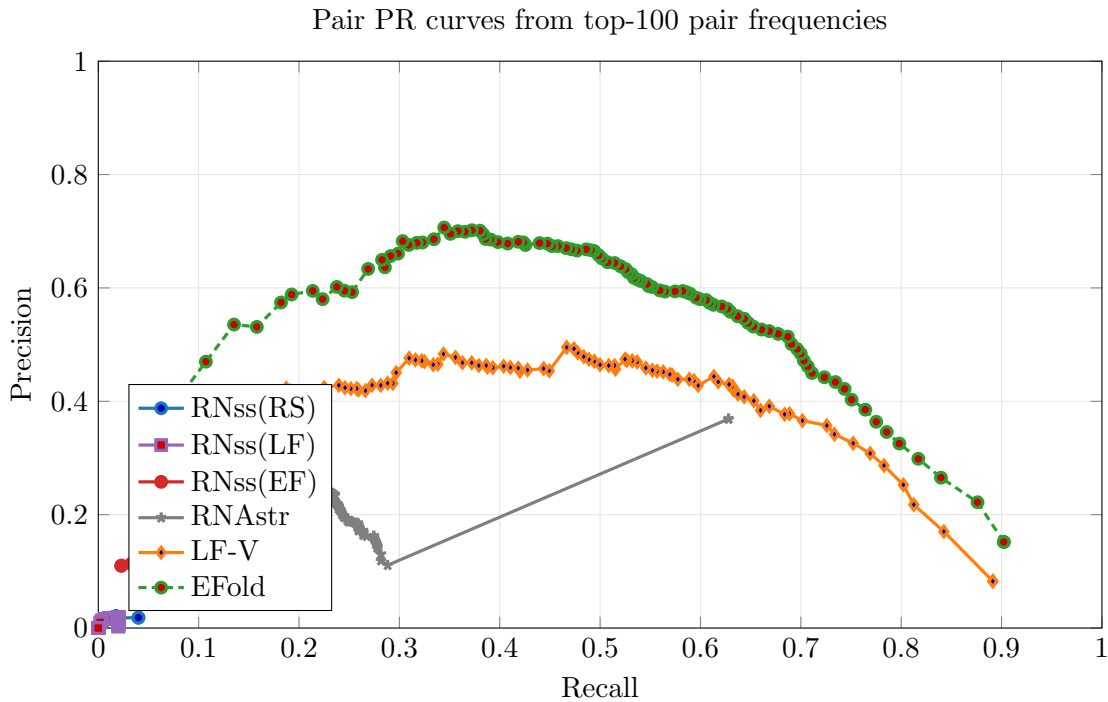
Precision/Recall Curves

Figure 4: Pair-level PR curves computed by thresholding pair probabilities estimated as *frequency across the top-100 predicted structures.*
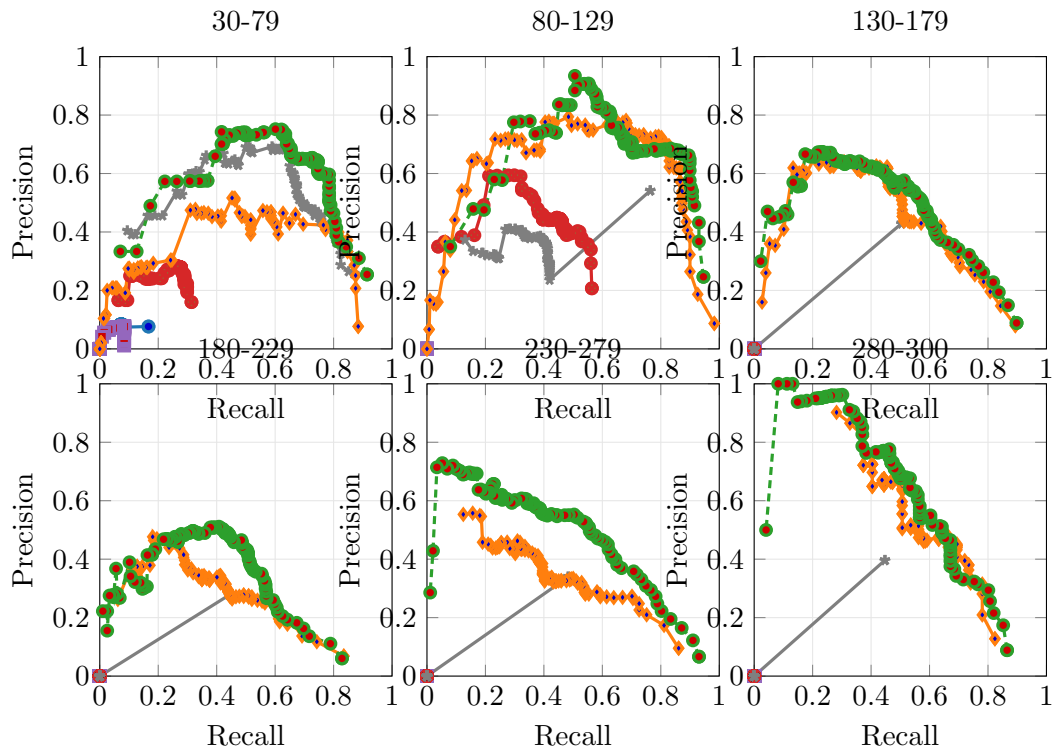
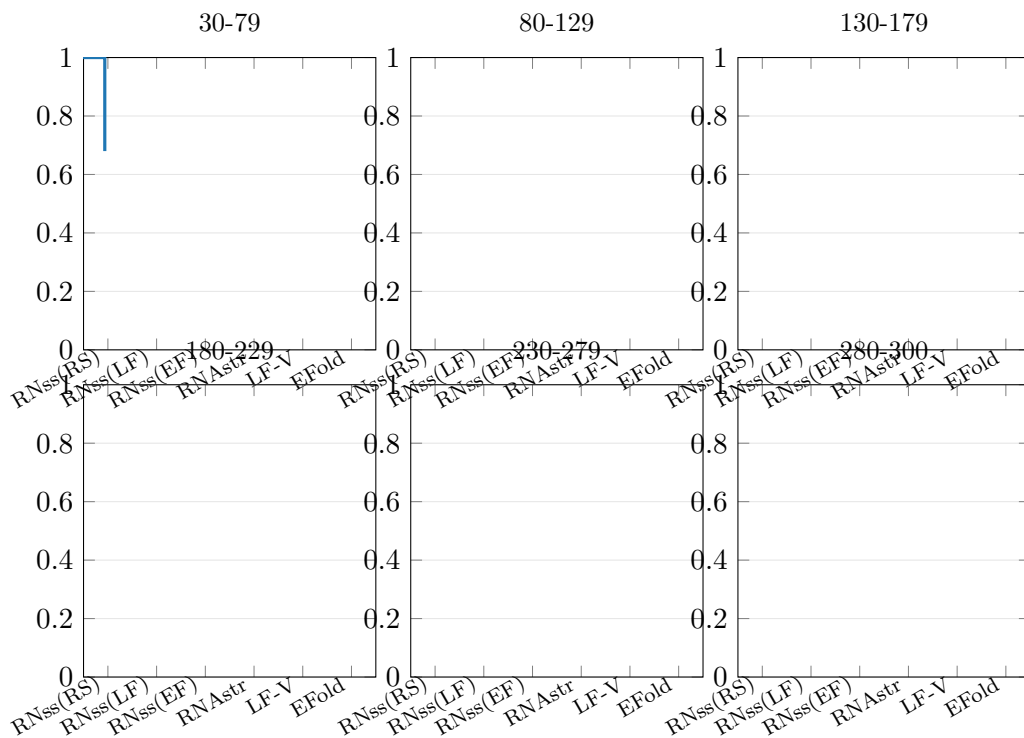Figure 5: PR curves by 50-nt length bucket.

Score Distributions

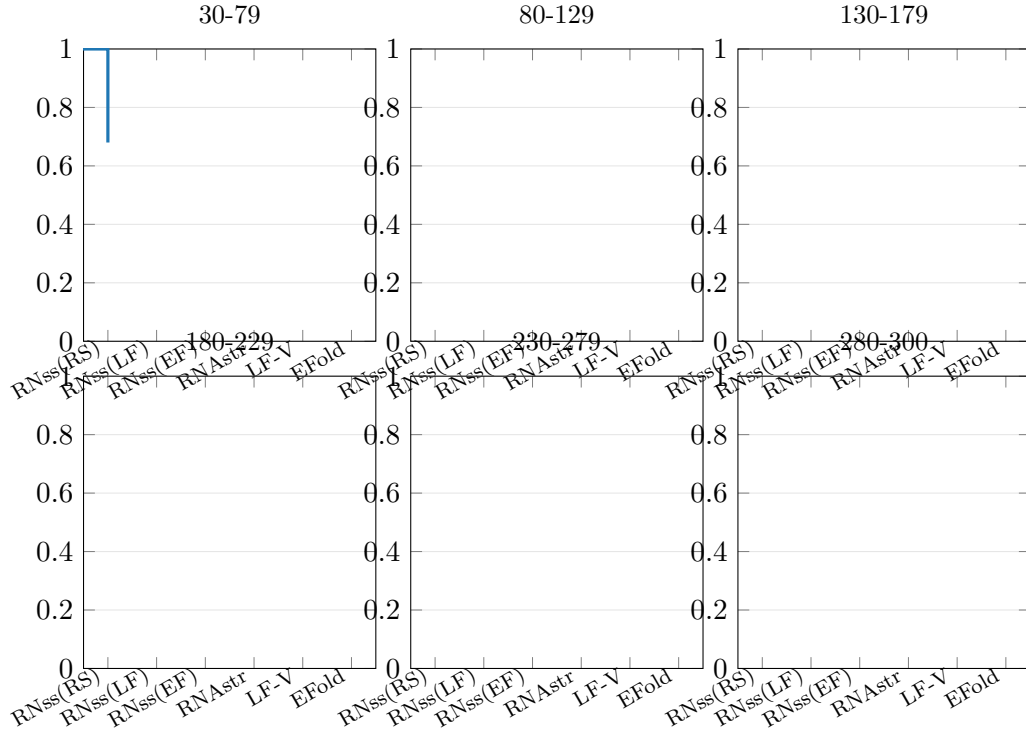Figure 6: F1@1 distributions by length bucket (boxplots; outliers hidden).

Figure 7: F1@100 distributions by length bucket (boxplots; outliers hidden).