APPENDIX A
EXPERIMENTAL RESULTS ON 250K ZINC DATASET

A. Effect of noise sampling methods on performance.

TABLE A.1
EFFECT OF NOISE SAMPLING METHODS ON PERFORMANCE.

Sampling method	Sampling size	Validity ↑	Uniqueness ↑	Novelty ↑	Diversity ↑
Normal	100	85 (85.00%)	80 (94.11%)	36 (45.00%)	0.93
	300	251 (83.60%)	216 (86.05%)	98 (45.00%)	0.93
	500	412 (82.40%)	335 (81.31%)	142 (42.39%)	0.93
	700	546 (78.00%)	416 (76.19%)	183 (43.99%)	0.93
	900	703 (78.33%)	501 (71.26%)	218 (43.51%)	0.93
	1000	779 (77.90%)	524 (67.26%)	226 (43.12%)	0.93
Uniform	100	68 (68.00%)	62 (91.17%)	27 (43.54%)	0.91
	300	209 (69.67%)	176 (84.21%)	74 (42.04%)	0.91
	500	343 (68.60%)	251 (73.17%)	101 (40.23%)	0.91
	700	441 (63.00%)	283 (64.17%)	113 (39.92%)	0.91
	900	564 (62.67%)	349 (61.87%)	136 (38.96%)	0.91
	1000	619 (61.90%)	366 (59.12%)	142 (28.79%)	0.91

B. Effect of training epochs on performance.

TABLE A.2
EFFECT OF TRAINING EPOCHS ON PERFORMANCE.

Epochs	Sampling size	Validity ↑	Uniqueness ↑	Novelty ↑	Diversity ↑
100	100	79 (79.00%)	74 (93.6%)	26 (35.13%)	0.89
	500	385 (77.00%)	301 (73.95%)	108 (35.88%)	0.89
	1000	723 (72.30%)	446 (61.69%)	161 (36.10%)	0.89
150	100	85 (85.00%)	80 (94.11%)	36 (45.00%)	0.93
	500	412 (82.40%)	335 (81.31%)	142 (42.39%)	0.93
	1000	779 (77.90%)	524 (67.26%)	226 (43.12%)	0.93
200	100	83 (83.00%)	76 (91.57%)	31(40.79%)	0.93
	500	403 (80.60%)	341 (84.62%)	139 (40.76%)	0.93
	1000	753 (75.30%)	552 (73.31%)	231 (41.85%)	0.93

C. Effect of training data volumes on performance.

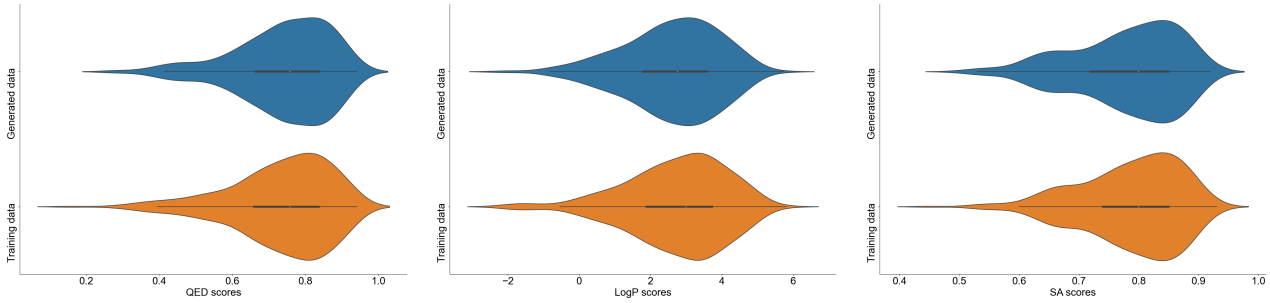
APPENDIX B
EXPERIMENTAL RESULTS ON 5K ZINC SUBSET

A. Effect of noise sampling methods on performance.

B. Effect of training epochs on performance.

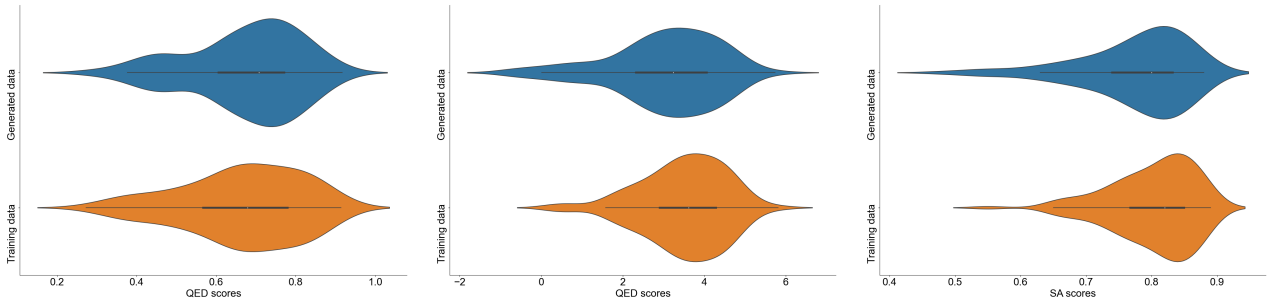
TABLE A.3
EFFECT OF TRAINING DATA VOLUMES ON PERFORMANCE.

Training data	Sampling size	Validity ↑	Uniqueness ↑	Novelty ↑	Diversity ↑
ZINC_25k	100	79 (79.00%)	69 (87.34%)	67 (97.10%)	0.94
	300	233 (77.66%)	193 (82.83%)	185 (95.85%)	0.94
	500	386 (77.20%)	294 (76.16%)	276 (93.88%)	0.94
	700	536 (76.57%)	387 (72.20%)	365 (94.32%)	0.94
	900	658 (73.11%)	425 (64.58%)	394 (92.71%)	0.94
	1000	723 (72.30%)	453 (62.65%)	412 (90.95%)	0.94
ZINC_250k	100	85 (85.00%)	80 (94.11%)	36 (45.00%)	0.93
	300	251 (83.60%)	216 (86.05%)	98 (45.00%)	0.93
	500	412 (82.40%)	335 (81.31%)	142 (42.39%)	0.93
	700	546 (78.00%)	416 (76.19%)	183 (43.99%)	0.93
	900	703 (78.33%)	501 (71.26%)	218 (43.51%)	0.93
	1000	779 (77.90%)	524 (67.26%)	226 (43.12%)	0.93



(a) Drug-likeness (QED) (b) Solubility (LogP) (c) Synthesizability (SA)

Fig. A.1. Distribution plots of QED, LogP, and SA scores on ZINC and generated data.



(a) Drug-likeness (QED) (b) Solubility (LogP) (c) Synthesizability (SA)

Fig. B.1. Distribution plots of QED, LogP, and SA scores on 5K ZINC and generated data.

TABLE B.1
EFFECT OF NOISE SAMPLING METHODS ON PERFORMANCE.

Sampling method	Sampling size	Validity ↑	Uniqueness ↑	Novelty ↑	Diversity ↑
Normal	100	62 (62.00%)	30 (48.38%)	30 (100.00%)	0.85
	300	198 (66.00%)	89 (44.94%)	89 (100.00%)	0.85
	500	350 (70.00%)	164 (46.85%)	160 (97.56%)	0.85
	700	446 (63.71%)	221 (49.55%)	215 (97.28%)	0.85
	900	606 (67.30%)	276 (45.54%)	253 (91.33%)	0.85
	1000	693 (69.30%)	313 (45.16%)	284 (90.73%)	0.85
Uniform	100	55 (55.00%)	25 (45.46%)	24 (96.00%)	0.83
	300	183 (67.30%)	81 (44.26%)	89 (97.53%)	0.83
	500	321 (64.20%)	146 (45.48%)	138 (94.52%)	0.83
	700	422 (60.28%)	191 (45.26%)	183 (95.81%)	0.83
	900	580 (64.44%)	241 (40.91%)	233 (96.68%)	0.83
	1000	669 (66.90%)	282 (42.15%)	270 (95.74%)	0.83

TABLE B.2
EFFECT OF TRAINING EPOCHS ON PERFORMANCE.

Epochs	Sampling size	Validity ↑	Uniqueness ↑	Novelty ↑	Diversity ↑
15	100	62 (62.00%)	30 (48.38%)	30 (100.00%)	0.85
	500	350 (70.00%)	164 (46.85%)	160 (97.56%)	0.85
	1000	693 (69.30%)	313 (45.16%)	284 (90.73%)	0.85
30	100	65 (65.00%)	32 (87.84%)	32 (98.46%)	0.84
	500	331 (66.20%)	214 (56.91%)	205 (95.79%)	0.84
	1000	667 (66.70%)	346 (47.07%)	329 (95.09%)	0.84
50	100	58 (58.00%)	30 (51.85%)	30 (100.00%)	0.84
	500	334 (66.80%)	159 (47.60%)	156 (98.11%)	0.83
	1000	675 (67.50%)	302 (44.74%)	284 (93.11%)	0.84