Table 1: Comparison between SemlaFlow and GeoRCG (Semla) across varying numbers of sampling steps. Evaluated on 5k samples on the GEOM-DRUG dataset, and with a single Nvidia 4090. Results for SemlaFlow are obtained from our own experiments. While classifier-free guidance approximately doubles the sampling time, GeoRCG still outperforms SemlaFlow even when using half the number of sampling *steps*—resulting in comparable sampling *time*.

# Steps	Method	Energy ↓	Energy/Atom ↓	Strain ↓	Strain/Atom ↓	Validity ↑	Atom-Stab. ↑	Mol.-Stab. ↑	Rep. Time ↓	Mol. Time ↓	Mol. Time w/o CFG ↓
		kcal· mol^{-1}	kcal· mol^{-1}	kcal· mol^{-1}	kcal· mol^{-1}	%	%	%	seconds	seconds	seconds
100	SemlaFlow	96.387	2.167	53.688	1.206	93.9	99.8	97.2	-	**610**	**610**
	GeoRCG (Semla)	**83.334**	**1.836**	**48.627**	**1.077**	**96.0**	**99.9**	**98.0**	97	1481	770
50	SemlaFlow	96.226	2.130	55.613	1.231	94.3	99.8	97.1	-	**310**	**310**
	GeoRCG (Semla)	**89.221**	**1.982**	**49.964**	**1.099**	**95.4**	**99.9**	**97.4**	97	690	380
20	SemlaFlow	106.388	2.359	68.080	1.516	94.1	**99.8**	**95.8**	-	**152**	**152**
	GeoRCG (Semla)	**95.984**	**2.119**	**65.038**	**1.449**	**94.6**	**99.8**	95.1	97	315	189

Table 2: Ablation study of GeoRCG (EDM) with CFG coefficient $w = 0.0$ and inverse temperature inv_T $= 1.0$, i.e., without classifier-free guidance and low-temperature sampling. Evaluated on 10k samples. The gray cells denotes the base molecule generator employed in GeoRCG (EDM). Even without these two techniques, GeoRCG consistently improves EDM by a significant margin and is highly competitive with advanced models.

	QM9			GEOM	
Model	**Mol.-Stab.** (%) ↑	**Valid**(%) ↑	**Valid & Unique**(%) ↑	**Atom-Stab.**(%) ↑	**Valid** (%)↑
EDM	82.0	91.90	90.70	81.3	92.6
GeoLDM	89.4	93.80	92.70	<u>84.4</u>	**99.3**
GeoBFN	<u>90.87</u>	95.31	**92.96**	**86.1**	91.66
GeoRCG (EDM) ($w = 1$, inv_T $= 1$)	**92.32**	**96.52**	<u>92.45</u>	84.3	<u>98.5</u>
GeoRCG (EDM) ($w = 0$, inv_T $= 1$)	90.2	<u>95.67</u>	92.65	<u>84.4</u>	97.8

Table 3: Energy and strain values for DRUG molecules generated by selected 3D-only models. Evaluated on 10k samples. All results are obtained by our own experiments. The gray cells denotes the base molecule generator employed in GeoRCG (EDM). Models marked with an asterisk (*) **are not directly comparable** to the methods below, as SemlaFlow is trained on the 5 lowest-energy conformations of DRUG, whereas the other methods are trained on the 30 lowest-energy conformations. GeoRCG consistently improves the base model by a large margin and outperforms GeoLDM.

Model	**Model type**	**QM9**				**DRUG**			
		Energy ↓	**Energy/Atom** ↓	**Strain** ↓	**Strain/Atom** ↓	**Energy** ↓	**Energy/Atom** ↓	**Strain** ↓	**Strain/Atom** ↓
SemlaFlow*	Flow (2D&3D)	—	—	—	—	96.387	2.167	53.688	1.206
GeoRCG (Semla)*		—	—	—	—	83.334	1.836	48.627	1.077
EDM	Diffusion	34.76	2.01	12.99	0.74	412.71	8.88	402.65	8.67
GeoLDM		<u>33.44</u>	<u>1.91</u>	<u>10.38</u>	<u>0.61</u>	<u>329.46</u>	<u>7.2</u>	<u>313.83</u>	<u>6.93</u>
GeoRCG (EDM)		**32.0**	**1.80**	**9.48**	**0.55**	**318.2**	**6.8**	**303.7**	**6.58**

Table 4: Conditional molecule generation on QM9. The metric used is the MSE between the target property value and the classifier-predicted value. SemlaFlow results are obtained by our own experiments. Results for SemlaFlow and GeoRCG (Semla) are calculated over 5k molecules. The gray cells and blue cells denote the base molecule generator employed in GeoRCG (EDM) and GeoRCG (Semla), respectively.

Methods \ Properties	Model type	α	$\varepsilon_{\text{LUMO}}$	$\Delta\epsilon$
QM9 (lower bound)	-	0.1	36	64
Random	-	9.01	1457	1470
N_atoms	-	3.86	813	866
EDM	Diffusion	2.76	584	655
GeoRCG (EDM)		**0.89(0.005)**	**290.8(3.1)**	**368(4.6)**
GeoLDM		2.37	522	587
GCDM		1.97	479	602
GeoBFN	Baysian Flow	2.34	516	577
EquiFM	Flow (3D-only)	2.41	530	591
GOAT		2.74	534	605
SemlaFlow	Flow (2D&3D)	1.63(0.004)	361.2(1.2)	429.5(4.1)
GeoRCG (Semla)		1.05(0.012)	340.1(0.8)	413.1(5.0)

Table 5: Reorganized Table 1 from the original paper to ensure more completeness. Quality comparison of unconditional molecular generation across 3D-only methods. The gray-background cells denote the base molecule generator employed in GeoRCG (EDM), and gray texts denote fewer-step versions of respective models.

			QM9				DRUG	
Methods \ Metrics	Model type	# Steps	Atom Sta (%) ↑	Mol Sta (%) ↑	Valid (%) ↑	Valid & Unique (%) ↑	Atom Sta (%) ↑	Valid (%) ↑
Data	-		99	95.2	97.7	97.7	86.5	99.9
G-Schnet	Autoregressive	-	95.7	68.1	85.5	80.3	-	-
GDM	Non-equivariant	1000	97	63.2	-	-	75	90.8
GDM-AUG		1000	97.6	71.6	90.4	89.5	77.7	91.8
GraphLDM		1000	97.2	70.5	83.6	82.7	76.2	97.2
GraphLDM-AUG		1000	97.9	78.7	90.5	89.5	79.6	98
EDM	Diffusion	50	97.0(0.1)	66.4(0.2)	-	-	-	-
		100	97.9(0.1)	69.8(0.2)	-	-	-	-
		500	98.5(0.1)	81.2(0.1)	-	-	-	-
		1000	98.7	82	91.9	90.7	81.3	92.6
EDM-Bridge		50	97.3	69.2	-	-	-	-
		100	97.9	72.3	-	-	-	-
		500	98.7	83.7	-	-	-	-
		1000	98.8	84.6	92	90.7	82.4	92.8
GeoLDM		1000	98.9(0.1)	89.4(0.5)	93.8(0.4)	92.7(0.5)	84.4	**99.3**
GCDM		1000	98.7(0.0)	85.7(0.4)	94.8(0.2)	93.3(0.0)	**89**	95.5
ENF	Flow (3D-only)	-	85	4.9	40.2	39.4	-	-
EquiFM		200	98.9(0.1)	88.3(0.3)	94.7(0.4)	**93.5(0.3)**	84.1	98.9
GOAT		90	98.4	84.1	90.9	89.99	81.8	96.0
GeoBFN	Bayesian Flow	50	98.28(0.1)	85.11(0.5)	92.27(0.4)	-	75.11	91.66
		100	98.64(0.1)	87.21(0.3)	93.03(0.3)	-	78.89	93.05
		500	98.78(0.8)	88.42(0.2)	93.35(0.2)	-	81.39	93.47
		1000	99.08(0.03)	90.87(0.1)	95.31(0.1)	92.96(0.1)	85.6	92.08
GeoRCG (EDM)	Two-stage	50	98.75(0.05)	89.08(0.52)	95.05(0.33)	-	81.44(0.1)	95.7(0.7)
		100	99.08(0.03)	91.85(0.34)	96.49(0.27)	-	83.02(0.06)	96.3(0.7)
		500	99.09(0.01)	91.89(0.24)	**96.57(0.12)**	-	84.03(0.37)	97.57(0.9)
		1000	**99.12(0.03)**	**92.32(0.06)**	96.52(0.2)	92.45(0.2)	84.3(0.12)	98.5(0.12)