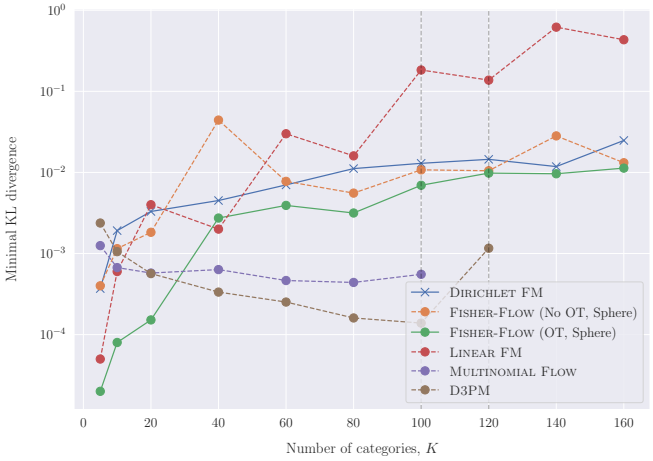

Figure 1: Extended toy experiment results: KL divergence against increasing number of categories, K with new baselines.

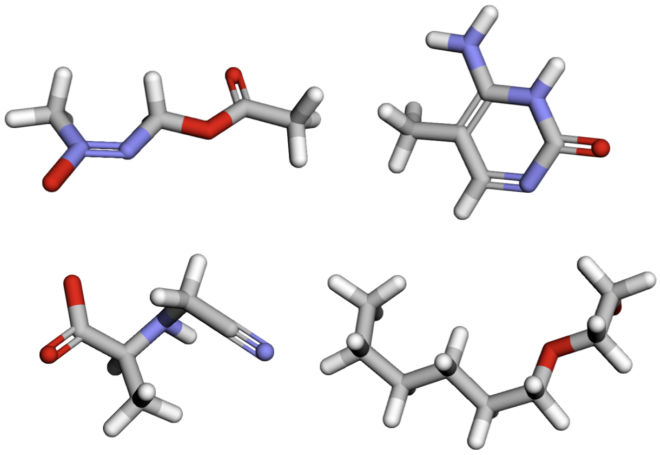

Figure 2: Generated molecules using Fisher-Flows on QM9.

Table 1: Updated results on the DNA datasets using an upper bound to the test perplexity and a Generated Perplexity metric which is the perplexity of generated samples as measured by an autoregressive model.

Dataset	Method	MSE	PPL (LL)	Gen-PPL
Promoter	Fisher-Flow (ours)	0.029 ± 0.001	$\leq 1.4 \pm 2.7$	1.001 ± 0.000
	Linear FM	0.056	≤ 1.381	1.001
	Dirichlet FM	0.034 ± 0.001	$\leq 1.978 \pm 0.006$	1.001 ± 0.000
	Random	—	—	4.454 ± 0.001
Enhancer Melanoma	Fisher-Flow (ours)	—	$\leq 1.4 \pm 0.1$	1.003 ± 0.0
	Linear FM	—	≤ 1.33	1.003
	Dirichlet FM	—	$\leq 2.25 \pm 0.01$	1.003 ± 0.0
	Random	—	—	4.452 ± 0.001
Enhancer FlyBrain	Fisher-Flow (ours)	—	$\leq 1.4 \pm 0.66$	1.009 ± 0.000
	Linear FM	—	≤ 1.35	1.008
	Dirichlet FM	—	$\leq 2.25 \pm 0.02$	1.008 ± 0.000
	Random	—	—	4.454 ± 0.001

Table 2: Test perplexities on the LM1B dataset. All baselines are taken from concurrent work MDLM Sahoo et al. [2024]. Best diffusion or flow-matching method is in bold font.

	Method	Parameters	PPL (↓)
Diffusion	BERT-Mouth	110M	≤ 142.89
	D3PM (absorb)	70M	≤ 77.50
	Diffusion-LM	80M	≤ 118.62
	DiffusionBert	110M	≤ 63.78
	SEDD (33B tokens)	110M	≤ 32.79
AR	Transformer (33B tokens)	110M	22.32
	Transformer (327B tokens)	110M	20.86
DM/FM	MDLM (33B tokens)	110M	≤ 27.04
	Fisher-Flow (33B tokens) (ours)	110M	$\leq$ **26.51**
DM/FM	MDLM (327B tokens)	110M	≤ 23.00
	Fisher-Flow (327B tokens) (ours)	110M	$\leq$ **22.42**

Table 3: New results on the QM9 dataset. Higher is better on all metrics. The baselines are taken from the cited papers.

Method	Atoms Stable (%)	Molecules Valid (%)	Molecules Stable (%)
Fisher-Flow (ours)	98.6	95.3	88.2
JODO [Huang et al., 2023]	99.4	98.9	98.7
EquiFM [Song et al., 2023]	99.4	94.4	93.2