A Supplementary Information

A.1 Ablation test: sampling D = D_{max}

D_{max} (Å)	Ratio of distorted:non-distorted molecules					
	1:20		1:50		1:100	
	RDKit Sanitisation, %	PoseBusters Pass Rate, %	RDKit Sanitisation, %	PoseBusters Pass Rate, %	RDKit Sanitisation, %	PoseBusters Pass Rate, %
0.1	91	46	**96**	**53**	81	26
0.25	81	2	81	2	72	4
0.5	49	0	28	0	38	0
1	41	0	46	0	29	0

Table 1: Performance comparison 100 molecules sampled with EDM trained conditionally on GEOM$_{\text{no h}}$ using a distortion factor, D, and sampled with D=D_{max}Å across various ratios of distorted:non-distorted molecules and maximum distortion values in angstrom.

A.2 Conditioning on internal energy

The eXtended Tight Binding (XTB) program [32] is a computational chemistry software package used for molecular modeling and simulations. It is based on the semi-empirical tight-binding approach, which approximates the electronic structure of molecules using a simplified set of parameters derived from quantum mechanics. XTB extends traditional tight-binding methods by incorporating additional empirical corrections to improve the accuracy of calculated properties. It is well-suited for studying large molecular systems where the computational cost of more accurate methods such as density functional theory (DFT) becomes prohibitive.

We use the default implementation of XTB to perform singlepoint energy calculations for each of the molecules, both distorted and original, for QM9, GEOM$_{\text{no h}}$, and our ZINC dataset. For the original GEOM dataset, the original conformers have energy annotations calculated with the same method, so we only carry out this calculation for the distorted molecules. We carry out this process using two different annotation types: a ‘distortion factor’, a quantity that represents the extent to which the coordinates have been altered, and an internal energy value, obtained by scoring both original and distorted conformers with the extended tight binding program (XTB) [32].

We’ve observed that for medium-sized, drug-like compounds, conditioning on a distance-based distortion factor results in improvements for RDKit sanitisation and PoseBusters tests. To investigate whether using a potentially more meaningful label—specifically, an internal energy value obtained using XTB—improves the conditioned models, we followed the previous distortion process. For one in every fifty molecules in each dataset, we generated a distorted version by distorting each atom’s coordinates by up to 0.25 Å and added these distorted molecules back to the dataset. These distorted molecules were then passed through an XTB single-point energy calculation. The same calculation was applied to all high-quality, non-distorted molecules in all datasets, excluding the original GEOM dataset, which already has energy annotations calculated with XTB. We then trained each of the models conditionally, and once trained, we sampled 100 molecules from each, this time enforcing D = E_{min}, where E_{min} is a fixed value for each dataset corresponding to the lowest internal energy annotation of any molecule in it. Then we once again assessed all 100 molecules using RDKit and PoseBusters (Table 2).

As observed when conditioning on distortion factor, both QM9 and the original GEOM dataset generated lower quality molecules when conditioning with internal energy was carried out. QM9 saw a further decrease in performance when using internal energy, while molecules generated by the model trained on GEOM saw a slight boost in RDKit performance but still exhibited a PoseBusters pass rate of 0%, ultimately being outperformed by the baseline molecules.

Conversely, conditionally training EDM on GEOM$_{\text{no h}}$ with internal energy resulted in increased performance compared to both the baseline and the conditioned model using the distance distortion factor. The RDKit and PoseBusters pass rates reached 98% and 84%, respectively. The model trained on ZINC, however, exhibited a decrease in both RDKit and PoseBusters pass rate. These pass rates were also lower than those of the molecules generated with the ZINC model trained conditionally on distortion factor.

Dataset		RDKit Sanitisation, %	PoseBusters Pass Rate, %
Baseline	QM9	95.2 (93.2–97)	67.6 (63.6–71.6)
	$\text{GEOM}_{\text{no h}}$	84.7 (82.5–86.9)	62.2 (58.3–66.1)
	ZINC	70.6 (67.8–73.3)	40.0 (37.0–43.0)
Conditional method (XTB)	QM9	57.1 (53.1-59.6)	19 (12–27)
	$\text{GEOM}_{\text{no h}}$	68 (59–77)	39 (30–49)
	ZINC	65 (56–74)	1 (0-3)

Table 2: Performance comparison of EDM trained on diverse molecular datasets using a baseline (the unconditional EDM), and using a conditioned model, for which the model is trained on an XTB internal energy estimate. 95% confidence intervals are shown in brackets.

When using the distance based distortion factor, we enforced D=0Å when sampling, as this represented molecules that had not undergone any distortion. The distortion factor, ranging from 0Å to D_{max}Å, provides a straightforward measure of how much the structure of a molecule has been altered. In contrast, internal energy values, although physically meaningful, are challenging to compare directly between different molecules. Each molecule can have a low internal energy corresponding to a high-quality conformer and a high one corresponding to a low-quality conformer, but the lowest energy conformers of some molecules may still have higher energy values than the highest energy annotations of others. This inconsistency likely contributed to the observed performance decreases when conditioning on internal energy.

A.3 Full PoseBusters outputs

Below we provide the full outputs for each model's molecules when assessed with PoseBusters.

Dataset	RDKit Sanitisation, %	Posebusters Pass Rate, %							
		All Atoms Connected	Bond Lengths	Bond Angles	Internal Steric Clash	Aromatic Ring Flatness	Double Bond Flatness	Internal Energy	All Tests Passed
No conditioning									
QM9	92.2 (90.5–93.8)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	99.9 (99.7–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	88.1 (85.9–90.1)	81.1 (78.7–83.5)
$\text{GEOM}_{\text{no h}}$	84.7 (82.5–86.9)	74.4 (71.7–77.1)	65.6 (62.5–68.7)	73.8 (70.8–76.7)	97.8 (96.7–98.7)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	75.2 (72.3–78.0)	62.2 (58.3–66.1)
ZINC	70.6 (67.8–73.3)	62.4 (59.4–65.3)	65.0 (61.5–68.4)	75.3 (72.2–78.5)	79.5 (76.6–82.4)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	78.5 (75.5–81.4)	40.0 (37.0–43.0)
Distortion factor conditioning									
QM9	65.0 (62.0–68.0)	84.0 (81.7–86.2)	96.6 (95.2–98.0)	95.8 (94.3–97.2)	98.9 (98.0–99.7)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	91.1 (88.8–93.2)	46.9 (43.9–50.0)
$\text{GEOM}_{\text{no h}}$	92.4 (90.7–94.0)	89.4 (87.5–91.3)	96.6 (95.5–97.8)	93.7 (92.1–95.2)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	87.7 (85.5–89.7)	68.8 (65.9–71.1)
ZINC	95.3 (93.8–96.6)	95.7 (94.3–96.9)	94.1 (92.5–95.6)	93.7 (92.1–95.2)	99.0 (98.4–99.6)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	89.4 (87.4–91.4)	74.5 (71.7–77.3)
Ablation tests									
1:20, D_{max} = 0.1Å	96 (92 –99)	93 (88 –98)	94 (89 –98)	92 (86 –97)	96 (92 –99)	96 (92 –99)	96 (92 –99)	82 (74 –89)	73 (64 –81)
1:20, D_{max} = 0.25Å	95 (90 –99)	73 (64 –82)	88 (81 –94)	90 (84 –95)	95 (90 –99)	95 (90 –99)	95 (90 –99)	82 (74 –89)	52 (42 –62)
1:20, D_{max} = 0.5Å	97 (93 –100)	95 (90 –99)	94 (89 –98)	89 (83 –95)	97 (93 –100)	97 (93 –100)	97 (93 –100)	88 (81 –94)	75 (66 –83)
1:20, D_{max} = 1Å	93 (88 –97)	88 (81 –94)	85 (78 –92)	85 (78 –92)	92 (86 –97)	93 (88 –97)	93 (88 –97)	76 (67 –84)	57 (47 –67)
1:50, D_{max} = 0.1Å	96 (92 –99)	93 (88 –97)	94 (89 –98)	94 (89 –98)	96 (92 –99)	96 (92 –99)	96 (92 –99)	86 (79 –92)	77 (69 –85)
1:50, D_{max} = 0.25Å	97 (92 –99)	91 (85 –96)	95 (90 –99)	94 (89 –98)	96 (92 –99)	96 (92 –99)	96 (92 –99)	90 (84 –96)	81 (73 –88)
1:50, D_{max} = 0.5Å	97 (93 –100)	90 (84 –95)	94 (89 –98)	96 (92 –99)	97 (93 –100)	97 (93 –100)	97 (93 –100)	91 (85 –96)	78 (70 –86)
1:50, D_{max} = 1Å	89.2 (78–97)	81.1 (68–92)	83.8 (70–95)	81.1 (68–92)	89.2 (78–97)	89.2 (78–97)	89.2 (78–97)	73 (57–86)	54.1 (38–70)
1:100, D_{max} = 0.1Å	96 (92 –99)	92 (86 –97)	95 (90 –99)	94 (89 –98)	96 (92 –99)	96 (92 –99)	96 (92 –99)	87 (80 –93)	77 (68 –85)
1:100, D_{max} = 0.25Å	96 (92 –99)	94 (89 –98)	95 (90 –99)	93 (88 –97)	96 (92 –99)	96 (92 –99)	96 (92 –99)	84 (77 –91)	77 (68 –85)
1:100, D_{max} = 0.5Å	95 (90 –99)	86 (79 –92)	94 (89 –98)	94 (89 –98)	95 (90 –99)	95 (90 –99)	95 (90 –99)	81 (73 –88)	68 (59 –77)
1:100, D_{max} = 1Å	62 (52 –71)	72 (63 –81)	42 (32 –52)	31 (22 –40)	56 (46 –66)	62 (52 –71)	62 (52 –71)	46 (36 –56)	8 (3 –14)
Energy conditioning									
QM9	57 (47–66)	40 (31–50)	56 (46–65)	57 (47–66)	56 (46–65)	57 (47–66)	57 (47–66)	53 (43–63)	19 (12–27)
$\text{GEOM}_{\text{no h}}$	68 (59–77)	67 (57–76)	68 (59–77)	68 (59–77)	68 (59–77)	68 (59–77)	68 (59–77)	63 (53–72)	39 (30–49)
ZINC	65 (56–74)	21 (13–29)	12 (6–19)	41 (31–51)	53 (43–63)	65 (56–74)	65 (56–74)	39 (30–49)	1 (0–3)
GCDM									
$\text{GEOM}_{\text{no h}}$ baseline	100.0 (100.0–100.0)	93.5 (91.9–95.0)	97.7 (96.7–98.6)	98.8 (98.1–99.4)	98.9 (98.2–99.5)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	86.2 (84.1–88.3)	77.8 (75.2–80.4)
$\text{GEOM}_{\text{no h}}$ conditioned	99.9 (99.7–100.0)	92.9 (91.2–94.4)	97.8 (96.9–98.6)	98.8 (98.1–99.4)	98.4 (97.6–99.1)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	88.6 (86.6–90.5)	79.7 (77.2–82.1)
ZINC baseline	56.3 (53.3–59.3)	63.7 (60.8–66.6)	95.3 (92.8–97.5)	97.5 (95.9–98.9)	91.2 (88.2–93.9)	100 (100–100)	99.7 (99.2–100)	88.2 (84.8–91.2)	40.8 (38–43.6)
ZINC conditioned	97.2 (95.8–98.4)	94.4 (92.5–96.1)	89.0 (86.3–91.5)	95.5 (93.7–97.1)	97.5 (96.0–98.7)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	82.0 (78.7–85.0)	66.5 (62.5–70.4)
MolFM									
$\text{GEOM}_{\text{no h}}$ baseline	100.0 (100.0–100.0)	93.5 (91.9–95.0)	97.7 (96.7–98.6)	98.8 (98.1–99.4)	98.9 (98.2–99.5)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	86.2 (84.1–88.3)	77.8 (75.2–80.4)
$\text{GEOM}_{\text{no h}}$ conditioned	100.0 (100.0–100.0)	92.8 (91.1–94.4)	97.5 (96.5–98.4)	98.7 (98.0–99.4)	98.3 (97.4–99.1)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	87.5 (85.4–89.5)	78.7 (76.2–81.2)
ZINC baseline	72.0 (69.2–74.8)	62.9 (59.9–65.9)	66.9 (63.5–70.4)	76.4 (73.2–79.4)	76.7 (73.5–79.7)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	79.2 (76.1–82.1)	42.3 (39.2–45.4)
ZINC conditioned	93.3 (91.6–94.9)	96.4 (95.1–97.5)	70.4 (67.4–73.5)	86.6 (84.3–88.9)	98.8 (97.9–99.5)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	76.6 (73.8–79.6)	45.9 (42.6–49.1

Table 3: Pass rates for each of the PoseBusters subtasks for each model trained.

A.4 Other metrics

In this section we assess structural diversity among the generated molecules using the MOSES framework [25], and check for presence of undesirable functional groups in each molecule using the REOS (Rapid Elimination of Swill[26, 27]) functionality in the `useful_rdkit_utils` toolkit[28]. This includes PAINS filters ('PAINS') [35], the SureChEMBL Non-MedChem Friendly SMARTS ('SureChEMBL') [36], the Bristol-Myers Squibb HTS Deck filters ('BMS') [37], the NIH MLSMR Excluded Functionality filters ('MLSMR') [38], the University of Dundee NTD Screening Library Filters ('Dundee') [39], filters of unwanted fragments derived by Inpharmatica Ltd. ('Inpharmatica') [40], the Pfizer lint filters ('LINT') [41], and the Glaxo Wellcome Hard filters ('Glaxo') [42].

Dataset	Diversity	REOS Filter Pass Rate, %							
		Glaxo	Dundee	BMS	PAINS	SureChEMBL	MLSMR	Inpharmatica	LINT
No conditioning									
QM9	0.81	70.7 (67.8–73.5)	58.3 (55.3–61.4)	59.5 (56.4–62.6)	95.7 (94.4–96.9)	69.4 (66.5–72.2)	57.7 (54.6–60.7)	65.9 (62.9–68.8)	67.9 (64.9–70.8)
$\text{GEOM}_{\text{no h}}$	0.68	34.4 (31.5–37.3)	10.3 (8.5–12.2)	12.5 (10.5–14.6)	84.6 (82.3–86.8)	28.6 (25.8–31.5)	14.9 (12.7–17.1)	24.1 (21.5–26.7)	20.9 (18.5–23.4)
ZINC	0.63	49.9 (44.8–50.7)	8.7 (6.7–10.0)	2.3 (1.3–3.1)	73.8 (67.8–73.3)	46.0 (41.0–47.0)	19.8 (16.7–21.3)	38.0 (33.5–39.3)	38.8 (34.3–40.0)
Conditioning on distortion factor									
QM9	0.72	44.1 (41.1–47.2)	34.3 (31.4–37.3)	42.6 (39.6–45.8)	86.4 (84.2–88.5)	41.6 (38.6–44.7)	37.7 (34.8–40.8)	41.7 (38.7–44.8)	40.1 (37.2–43.2)
$\text{GEOM}_{\text{no h}}$	0.74	57.7 (54.6–60.8)	23.4 (20.8–26.0)	21.5 (19.0–24.1)	92.4 (90.7–94.0)	49.5 (46.4–52.6)	33.7 (30.8–36.7)	49.3 (46.2–52.4)	50.7 (47.6–53.8)
ZINC	0.77	77.2 (76.9–82.0)	12.9 (11.2–15.4)	4.8 (3.6–6.4)	92.6 (93.8–96.6)	70.6 (69.8–75.4)	23.7 (21.7–27.1)	56.8 (55.2–61.4)	54.4 (52.8–59.1)
GCDM									
$\text{GEOM}_{\text{no h}}$ baseline	0.86	77.5 (74.9–80.0)	29.2 (26.4–32.0)	29.1 (26.3–32.0)	99.6 (99.2–99.9)	65.5 (62.5–68.4)	46.0 (42.9–49.1)	58.5 (55.4–61.6)	60.1 (57.1–63.1)
$\text{GEOM}_{\text{no h}}$ conditioned	0.84	73.5 (66.7–80.2)	37.7 (30.2–45.1)	29.0 (22.2–35.8)	99.4 (98.1–100.0)	68.5 (61.1–75.3)	38.3 (30.9–45.7)	56.2 (48.8–63.6)	59.3 (51.9–66.7)
ZINC baseline	0.61	34.1 (31.2–37.0)	17.6 (15.2–20.0)	26.6 (23.9–29.3)	36.2 (33.2–39.2)	27.3 (24.6–30.1)	9.7 (7.9–11.6)	27.9 (25.2–30.7)	23.5 (20.9–26.1)
ZINC conditioned	0.81	81.1 (77.9–84.2)	18.7 (15.6–21.9)	11.0 (8.6–13.7)	97.2 (95.8–98.4)	71.8 (68.1–75.5)	30.3 (26.6–34.0)	62.9 (58.8–66.7)	58.1 (54.1–62.0)
MolFM									
$\text{GEOM}_{\text{no h}}$ baseline	0.83	66.9 (62.7–71.0)	24.7 (21.0–28.4)	24.3 (20.6–28.0)	98.6 (97.5–99.6)	57.6 (53.3–62.0)	35.9 (31.8–40.0)	55.5 (51.2–59.8)	58.4 (54.1–62.7)
$\text{GEOM}_{\text{no h}}$ conditioned	0.77	58.3 (55.2–61.4)	20.7 (18.2–23.3)	19.6 (17.2–22.1)	94.5 (93.0–95.9)	50.5 (47.4–53.5)	29.2 (26.4–32.0)	50.8 (47.6–53.9)	50.2 (47.1–53.3)
ZINC baseline	0.64	51.3 (48.2–54.4)	7.6 (6.0–9.3)	2.7 (1.8–3.7)	72.0 (69.2–74.8)	46.5 (43.4–49.6)	17.4 (15.1–19.8)	38.4 (35.3–41.4)	40.2 (37.3–43.2)
ZINC conditioned	0.74	72.3 (69.4–75.1)	14.5 (12.3–16.9)	6.6 (5.0–8.3)	93.3 (91.6–94.9)	68.7 (65.7–71.6)	23.6 (20.9–26.4)	57.8 (54.5–61.0)	59.4 (56.2–62.5)

Table 4: Pass rates for REOS tests assessing baseline and conditional EDM, GCDM and MolFM trained on QM9, $\text{GEOM}_{\text{no h}}$ and ZINC.