

On-Demand Reverse Design of Polymers with PolyTAO

Haoke Qiu, ^{ab} Zhao-Yan Sun ^{*ab}

^a State Key Laboratory of Polymer Physics and Chemistry & Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, China.

^b School of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei 230026, China

E-mail: zysun@ciac.ac.cn (Z.Y.S.)

Supplementary Information

Table of contents

S1. Fundamental polymer properties selected	3
S2. Structurally diverse polymers generation	4
S3. Proxy model for evaluation the properties of the generated polymers	5

S1. Fundamental polymer properties selected

Table S1. The detailed list and description of the chosen Physicochemical properties of polymers.

Type	Properties of polymers	Description
Microscopic interactions-related	<i>NumHAcceptors</i>	The number of hydrogen bond acceptors
	<i>NumHDonors</i>	The number of hydrogen bond donors
	<i>NHOHCount</i>	The number of NHs or OHs
	<i>NOCOUNT</i>	The number of nitrogens and oxygens
Structure-related	<i>NumAliphaticCarbocycles</i>	The number of aliphatic (containing at least one non-aromatic bond) carbocycles
	<i>NumAliphaticHeterocycles</i>	The number of aliphatic (containing at least one non-aromatic bond) heterocycles
	<i>NumAliphaticRings</i>	The number of aliphatic (containing at least one non-aromatic bond) rings
	<i>NumAromaticCarbocycles</i>	The number of aromatic carbocycles
	<i>NumAromaticHeterocycles</i>	The number of aromatic heterocycles
	<i>NumAromaticRings</i>	The number of aromatic rings
	<i>RingCount</i>	The number of rings
Atom-related	<i>NumRotatableBonds</i>	The number of rotatable bonds
	<i>NumHeteroatoms</i>	The number of heteroatoms
	<i>HeavyAtomCount</i>	The number of heavy atoms
	<i>MolWt</i>	The exact molecular weight of the molecule

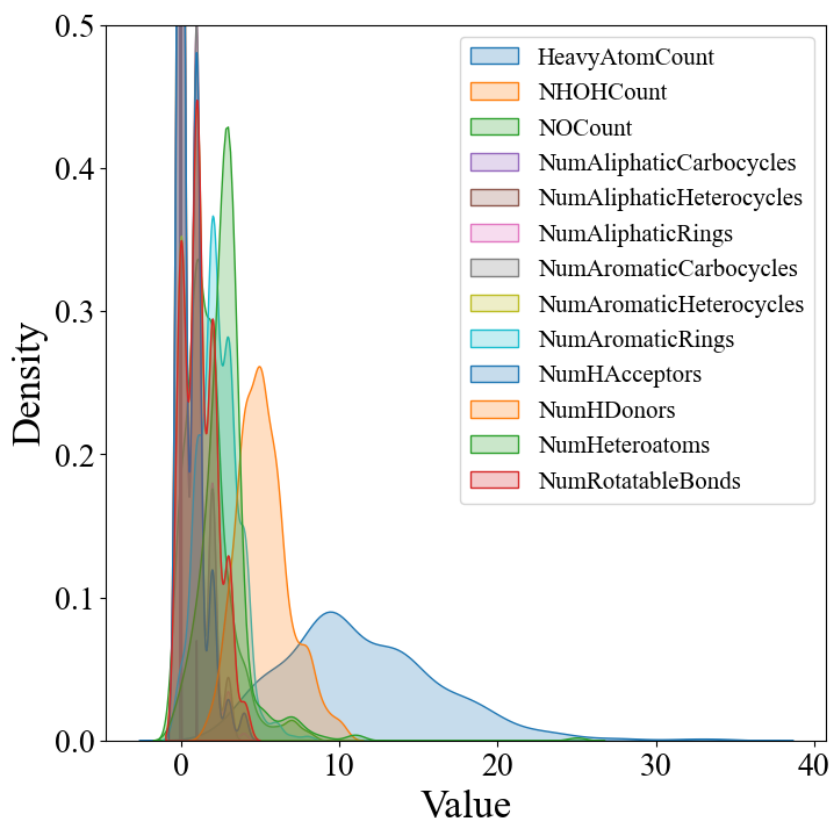


Figure S1. Value distribution of each property.

S2. Structurally diverse polymers generation

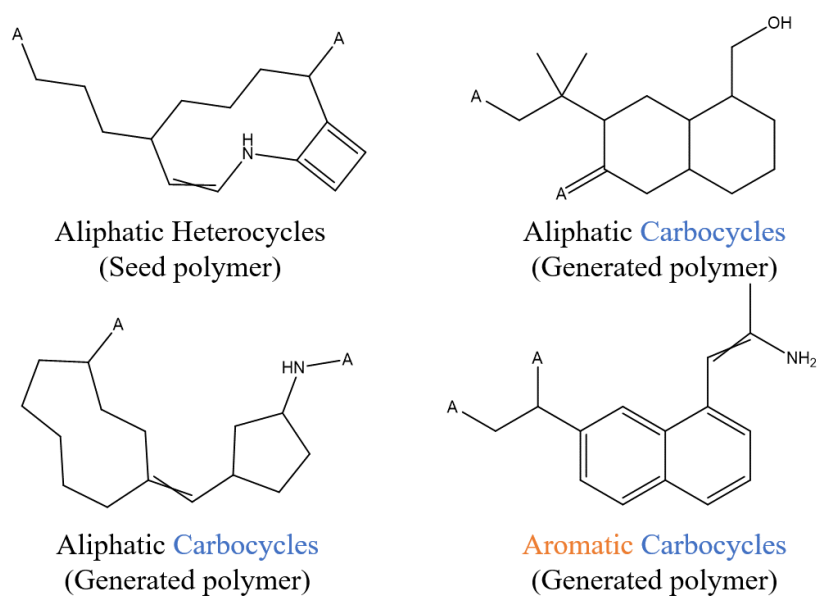


Figure S2. Polymer with structural diverse structures generation. Different font colors are used to highlight variations in the chain structure to . This indicates that the pretrained model can generate structurally diverse polymers.

S3. Proxy model for evaluation the properties of the generated polymers

We trained proxy models for evaluation the properties of the generated polymers using graph neural networks on ten collected publicly available datasets: atomization energy, band gap bulk, band gap chain, band gap crystal, crystallization tendency, dielectric constant, electron affinity, ionization energy, refractive index and glass transition temperature (Tg)¹⁻⁴. We configured the graph neural networks according to previous work. Each proxy model was trained five times to eliminate randomness. The performance of each property's proxy model is shown in the table below.

Table S2. Performance of each proxy model (ordered by R²).

Proxy model	# Data	R ² on test set
atomization energy	390	0.96 (+/- 0.01)
electron affinity	368	0.91 (+/- 0.01)
band gap chain	3380	0.90 (+/- 0.01)
band gap bulk	561	0.90 (+/- 0.01)
Tg	314	0.84 (+/- 0.02) ³
band gap crystal	236	0.83 (+/- 0.04)
ionization energy	370	0.81 (+/- 0.04)
refractive index	382	0.77 (+/- 0.02)
dielectric constant	382	0.78 (+/- 0.04)
crystallization tendency	432	0.47 (+/- 0.09)

We aim to use proxy models with sufficiently high accuracy to evaluate the properties of generated polymers, while also ensuring that the dataset has as much data as possible for fine-tuning the large language model. Taking these factors into account, we choose the following three datasets for downstream task validation: atomization energy, electron affinity, band gap chain (denoted as “band gap” in the main text).

Reference:

1. Kuenneth, C. *et al.* Polymer informatics with multi-task learning. *Patterns* **2**, 100238 (2021).
2. Kamal, D. *et al.* Novel high voltage polymer insulators using computational and data-driven techniques. *J. Chem. Phys.* **154**, 174906 (2021).
3. Bicerano, J. Prediction of polymer properties; CRC Press (2002).
4. Qiu, H., Wang, J., Qiu, X., Dai, X. & Sun, Z.-Y. Heat-resistant polymer discovery by utilizing interpretable graph neural network with small data. *Macromolecules* **57**, 3515–3528 (2024).