Periodicity-aware deep learning for polymers

Yuhui Wu[1,2,4], Cong Wang[1,2,4], Xintian Shen[1,2], Tianyi Zhang[1,2], Peng Zhang[1,2,3], Jian Ji[1,2,3]**

[1] MOE Key Laboratory of Macromolecule Synthesis and Functionalization, Department of Polymer Science and Engineering, Zhejiang University, Hangzhou, 310027, PR China

[2] International Research Center for X Polymers, International Campus, Zhejiang University, Haining, 314400, PR China

[3] State Key Laboratory of Transvascular Implantation Devices, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, P. R. China

[4] These authors contributed equally: Yuhui Wu, Cong Wang

** The corresponding authors (zhangp7@zju.edu.cn, jijian@zju.edu.cn).*

Abstract

Deep learning has revolutionized chemical research by accelerating the discovery of new substances and enhancing the understanding of complex chemical systems. However, polymer chemistry, one of the most active branches of chemistry, has yet to establish a unified deep learning framework due to the complexity of polymer structures. Existing self-supervised learning methods for polymers simplify them into repeating units and neglect their inherent periodicity, thereby limiting the models' ability to generalize across various tasks. To address this challenge, we herein propose a periodicity-aware deep learning framework for polymers (named PerioGT). In pre-training, a chemical knowledge-driven periodicity prior is constructed and incorporated into the model through contrastive learning. Then, periodicity prompts are learned in fine-tuning based on the periodicity prior to better leveraging the knowledge acquired in pre-training. Additionally, a novel graph augmentation strategy is employed for polymers, which integrates additional conditions via virtual nodes to effectively model complex chemical interactions. PerioGT achieves state-of-the-art performance on 12 downstream tasks. Moreover, wet-lab experiments in antimicrobial polymer discovery highlight PerioGT's potential in the real world, identifying two polymers with potent antimicrobial properties. All the results demonstrate that introducing the periodicity prior of polymers effectively improves the model performance.

Introduction

Deep learning (DL) has already had a transformative impact on chemical research due to its powerful ability to model complex relationships within chemical data. Significant breakthroughs have been recently achieved in biochemistry driven by DL, such as protein structure prediction[1-3], de novo protein design[4,5], and RNA sequence design[6,7]. In the field of organic chemistry, remarkable advancements have been made in molecular property prediction[8-10] and reaction condition optimization[11] enabled by DL. In inorganic chemistry, researchers have made notable progress in predicting crystal structures[12,13]. These advancements have profoundly enhanced our understanding of complex chemical systems and accelerated the discovery of new substances. Within this revolution, self-supervised learning (SSL) stands out[14-17]. SSL methods can extract meaningful representations by mining massive unlabeled data. The pre-training and fine-tuning paradigm of SSL has the potential to alleviate the scarcity of labeled data in chemistry.

Despite being one of the most active branches of chemistry, polymer chemistry still lacks a

universally applicable deep learning approach attributed to the inherent complexity. Polymers are macromolecules consisting of several to thousands of repeating units (RUs) linked by covalent bonds. This structural characteristic endows them with unique physicochemical properties that are significantly different from small molecules. Consequently, polymers became the most widely used materials. In addition to the chemical structure of RUs, many other variables, including the degree of polymerization, polydispersity index, crosslinking degree, copolymerization, can regulate the properties of polymers. The complexity of polymers provides a wide space for chemists to design polymers with various properties, yet it poses considerable challenges to their modeling.

Although there have been some pioneering studies that have successfully applied machine learning to polymer design[18-22], the severe scarcity of labeled data remains one of the major difficulties due to the complexity of the structure. Therefore, developing polymer SSL models is an effective approach. In the existing SSL models, polymers are simplified into RUs derived from monomers and represented as SMILES, akin to small molecules[23,24]. Language models are then trained through Masked Language Modeling (MLM) with the processed unlabeled datasets, followed by fine-tuning for downstream tasks. However, these methods treat polymers as small molecules and do not consider their periodic characteristics, which is the most significant difference between polymers and small molecules. The simplification makes it difficult for the model to capture the periodic structure of polymers. Therefore, incorporating periodicity as a prior is essential for mitigating issues such as data scarcity and limited robustness.

Attempts have been made in some supervised methods by adding an edge between the head and tail of the RU graph, thereby simulating the periodic structure[25,26]. Nevertheless, these methods still fail to adequately model the periodicity, leading to changes in polymer topology that introduce ambiguity and negatively impact model generalization (refer to Supplementary Fig. 4 for details).

Moreover, chemical structure alone cannot fully represent polymers in some cases, as their properties are regulated by multiple factors mentioned above. An ideal framework needs to possess a mechanism for integrating supplementary conditions. In some studies, supplementary conditions such as molecular weight or external factors are treated as a special token and concatenated with the SMILES[24]. This approach implicitly encodes the relationships between conditions and polymer structures, relying on the model to deduce these from sequence patterns. Such indirect representation may constrain the model's understanding of complex chemical principles. Therefore, it is crucial to develop a unified deep learning framework that explicitly models the behavior of polymers under various conditions, thereby accurately capturing complex chemical interactions.

Motivated by the challenges above, we propose a periodicity-aware deep learning model for polymers (PerioGT). Our contributions are as follows: (1) We introduce a chemical knowledge-driven sampling module, named Periodicity Augmentation (PA), serving as a periodic prior. (2) We propose a PA-based contrastive learning pre-training, which integrates periodic information as a prior for polymers, enhancing the model's ability to recognize and encode periodic patterns. (3) Periodicity prompt guided fine-tuning (PGFT) strategy is employed, introducing periodic prompts based on PA in fine-tuning to better leverage the knowledge learned in pre-training. (4) To establish a unified deep learning framework for polymers, we propose a modular graph augmentation strategy (named PolyGraph), explicitly modeling the relationships between additional information and polymers. (5) As a case study, the proposed model is further applied to the screening of antimicrobial polymers, and we successfully screened 25 novel polymers with antimicrobial activity, two of which show excellent antimicrobial activity against methicillin-resistant *Staphylococcus aureus* (MRSA).

Results and discussion

An overview of PerioGT

In this work, we propose a unified, periodicity-aware deep learning framework for polymers, PerioGT. The framework consists of three components: (1) PolyGraph construction, (2) Periodicity-aware SSL, and (3) Periodicity prompt guided fine-tuning (PGFT). The overview of PerioGT is shown as Fig. 1.

PolyGraph construction. Polymers have a complex multilevel structure, and the chemical structure of RUs alone is not sufficient to represent its structure. As a result, standard descriptors for small molecules are insufficient to accurately represent polymers.

To this end, we provide a flexible mechanism that can adapt to more complex cases. Additional structural information of polymers, such as the degree of polymerization, polydispersity index, crosslinking degree, and end groups, if available, can be incorporated as a virtual node $\hat{v}_{\text{add}}$ into the graph[27,28], thereby constructing a PolyGraph as shown in Fig. 1c. For the case of more complicated copolymers, a global virtual node $\hat{v}_{\text{glob}}$ is introduced, connecting to the nodes of all components. Additionally, local virtual nodes $\{\hat{v}_i^{\text{local}}\}$ are introduced, each connecting to the nodes of their respective components. Global features, such as molecular weight and test conditions, can be incorporated into the global virtual node $\hat{v}_{\text{glob}}$. Similarly, local features, such as the content of each component, can be added to the local virtual node $\{\hat{v}_i^{\text{local}}\}$. Moreover, we consider that representing polymers only with RUs may pose certain risks. Even if the two RUs are similar, these differences may accumulate during polymerization, eventually leading to large differences in properties (Supplementary Fig. 5). Therefore, another virtual node $\hat{v}_{\text{expt}}$ with physicochemical properties of polymers is introduced to guide the model in better-extracting polymer representations. More details on PolyGraph construction can be found in the Methods section. We suppose that modular and explicit modeling of PolyGraph, aligned with chemical intuition, can facilitate the model's ability to capture complex chemical interactions.

Periodicity-aware SSL. Existing SSL methods simplify polymers to minimal repeating units (MRUs) derived from their monomer and add a special token [*] to the SMILES of these MRUs to indicate the repeating sites, followed by MLM[23,24]. These models ignore that polymers are periodically arranged by RUs, which is one of the most remarkable characteristics of polymers. For instance, the SMILES "**NCCCCCC(=O)**" and "**C(=O)NCCCCC**" both denote nylon-6. However, the representations generated by these SSL models may not converge in the latent space. The inconsistency indicates that the conventional representation learning strategies applicable to small molecules are inadequate for enabling the model to learn the periodic patterns of polymers.

To explicitly model the periodicity and take advantage of the large amount of unlabeled data, we propose a contrastive learning strategy based on Periodicity Augmentation (PA). As shown in Fig. 1a, for a polymer chain, we construct a sampling box sized at N times the number of backbone atoms in an MRU. By arbitrarily sliding this sampling box along the polymer backbone, various augmented fragments can be obtained, all of which can be deemed as RU of the polymer. PA, as a prior of periodicity, provides different views of the same polymer, including different orders of arrangement and different degrees of polymerization. Thus, augmented fragments via PA are taken as positive pairs, and fragments sampled from different polymers are taken as negative pairs. Our approach, which enforces the model to align fragments sampled from the same polymer, aims to

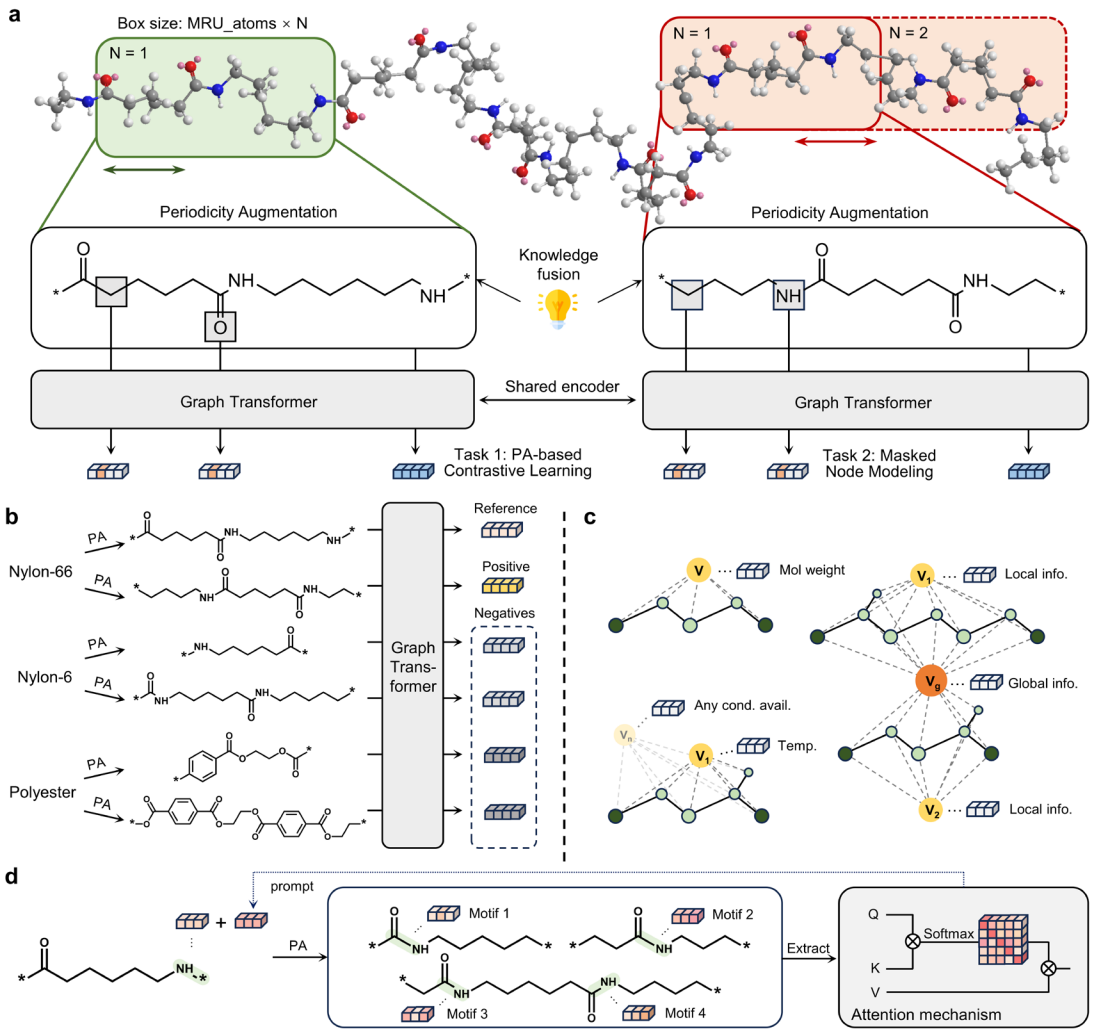

Fig. 1 | Overview of PerioGT. a, PerioGT pre-training. A sampling box is constructed with a size N times the number of main chain atoms in the minimal repeating unit. Sampling is performed by sliding the sampling box along the polymer backbone. Based on the sampled fragments, PolyGraphs are constructed, and two random masking operators are applied to each PolyGraph. A graph transformer encoder is used to extract node-level representations, and graph-level representations are obtained with the addition of a readout operator. Graph-level and node-level representations are then used for PA-based contrastive learning and masked node modelling, respectively. **b**, Illustration of PA-based contrastive learning. Given the reference fragment, the positive fragment is randomly sampled from the same polymer through PA while negative fragments are sampled from other polymers. **c**, PolyGraph construction. Any additional information available of polymers is fused through virtual nodes. These virtual nodes are directly connected to all real nodes by special edges. For copolymers, a global virtual node is added and connected to all real nodes. Additionally, n local virtual nodes are added, each connected to nodes corresponding to specific components, where n represents the number of copolymer components. Global and local information are introduced as features of the respective virtual nodes. **d**, Periodicity prompt guided fine-tuning. Given a polymer, PA is employed to obtain a set of augmentations containing M RUs. A feature extractor is employed to extract representations of the corresponding structures in the PA. Then these representations are integrated through an attention mechanism and added as a prompt to the initial node features.

enhance the model's ability to recognize the periodicity of polymers and guide it to capture the long-range interactions within polymer chains. Besides, previous SSL models for polymers are based on MLM. However, unlike long-distance dependencies common in natural language, the chemical properties of an atom are primarily determined by its adjacent atoms and bonds. This distinction suggests that MLM can only capture local features of polymers. We believe that PA-based contrastive learning not only introduces a periodicity prior, but also compensates for the limitations of MLM in capturing global features.

Specifically, as illustrated in Fig. 1b, given a batch containing N polymers, random augmentations are performed on each polymer via PA, resulting in $2N$ RUs. These RUs are then transformed into two sets of PolyGraphs $\{\hat{G}_i\}_{i=1}^{N}$, $\{\hat{G}_i'\}_{i=1}^{N}$. To prevent the model from over-relying on obvious differences between oligomers and to promote the learning of essential features, each PolyGraph undergoes random masking. Next, PolyGraphs derived from the same polymer are treated as positive pairs, whereas the remaining $2(N\text{-}1)$ graphs are considered negative pairs. We use a graph encoder $f(\cdot)$ with a readout operator to extract graph features followed by a non-linear projector $g_{\mathrm{PACL}}(\cdot)$ to map graph representations into a space for contrastive learning, generating latent vectors $\{\boldsymbol{z}_i\}_{i=1}^{N}$ and $\{\boldsymbol{z}_i'\}_{i=1}^{N}$. Contrastive Loss is applied to these latent vectors to maximize the agreement between positive pairs while minimizing the agreement between negative pairs. Another non-linear projector $g_{\mathrm{MNM}}(\cdot)$ maps the perturbed nodes $\mathcal{M}$ into a space, where Cross Entropy Loss is applied to minimize the prediction error of the masked nodes. Meanwhile, previous studies on small molecules have shown that the reconstruction of physicochemical properties from molecular graphs has a positive effect[29,30], so the reconstruction of partially masked virtual nodes is also conducted as one of the training objectives. More information on SSL is provided in the Methods section.

Periodicity prompt generation. Most existing pre-trained models follow the "pre-training, fine-tuning" strategy, which utilizes the weights of the pre-trained model as initialization and finetunes the weights for a specific downstream task[31]. However, the misalignment between the objectives of the pre-training task and the downstream task leads to sub-optimal performance[32]. Recently, prompt tuning has shown great promise in natural language processing and computer vision[33,34]. This method modifies the input data to prompt the model to utilize the knowledge learned in pre-training. Due to the complexity of graph data structures, systematic studies on prompt generation approaches for graphs are still lacking. Pioneering works have explored adding a learnable vector or external information to node features to improve alignment with pre-training objectives[35-37].

Inspired by this, we provide an option that aggregates the equivalent motifs in different RUs as a prompt in fine-tuning. Periodicity prompt emphasizes the periodic structure of the polymer, guiding the model to capture the environmental diversity of atoms at specific locations of the polymer. Given a polymer fragment $\hat{G}$, PA is employed to obtain a set $\{\hat{G}_i\}_{i\in[1,N_{\mathrm{aug}}]}$ containing N_{aug} RUs. As shown in Fig. 1d, for any triplet $(\mathrm{atom}_i, \mathrm{bond}_j, \mathrm{atom}_k)$ in $\hat{G}$, features of corresponding structures from $\{\hat{G}_i\}_{i\in[1,N_{\mathrm{aug}}]}$ are integrated through an attention mechanism and added as a prompt to the initial node features. More details can be found in the Methods section.

PerioGT enhances the performance of polymer property prediction

Polymer property prediction tasks lack benchmark datasets. To effectively evaluate and compare model performance, we utilized datasets recommended by previous studies. For more detailed

Table 1 | Test performance of different models on eight downstream tasks

Dataset	Tg	Mt	Egc	Density	Egb	Eps	Eea	Nc
Number of polymers	**7170**	**3651**	**3380**	**1693**	**561**	**382**	**368**	**382**
RF	40.712±0.108	57.386±0.438	0.539±0.004	0.0906±0.0003	0.579±0.011	0.563±0.010	0.409±0.005	0.1103±0.0153
MPNN[38]	36.213±0.153	58.722±1.095	0.509±0.025	0.0957±0.0083	0.535±0.027	0.474±0.001	0.332±0.016	0.0932±0.0023
GAT[39]	39.810±1.211	57.930±1.106	0.488±0.018	0.0948±0.0081	0.553±0.026	0.630±0.082	0.382±0.074	0.1067±0.0006
Graphormer[40]	37.201±0.532	55.297±2.054	0.500±0.024	0.1176±0.0046	0.546±0.004	0.583±0.049	0.434±0.105	0.1257±0.0125
LiGhT[41]	35.534±0.529	57.457±0.562	0.469±0.020	0.0934±0.0067	0.486±0.018	0.586±0.045	0.378±0.016	0.1223±0.0236
GraphGPS[42]	34.176±0.169	50.620±0.823	0.464±0.018	0.1003±0.0116	0.493±0.062	0.606±0.086	0.323±0.017	0.1120±0.0121
Arora et al.[22]	42.256±0.360	59.109±0.242	0.575±0.000	0.1099±0.0005	0.565±0.009	0.597±0.006	0.516±0.004	0.1098±0.0023
Antoniku et al.[25]	35.393±0.313	59.898±0.662	0.487±0.013	0.0915±0.0038	0.508±0.005	0.474±0.011	0.316±0.017	0.0993±0.0035
Aldeghi et al.[26]	35.470±0.455	55.637±0.415	0.456±0.006	0.0901±0.0033	0.442±0.009	0.510±0.019	0.274±0.006	0.0869±0.0049
polyBERT[23]	35.110±1.059	57.024±0.630	0.474±0.026	0.1020±0.0034	0.538±0.021	0.491±0.035	0.327±0.041	0.0970±0.0026
TransPolymer[24]	32.765±1.051	51.267±1.399	0.437±0.005	0.0881±0.0007	0.570±0.035	0.573±0.029	0.320±0.005	0.1301±0.0175
PerioGT	**30.840±0.221**	**45.439±0.397**	**0.399±0.003**	**0.0794±0.0018**	**0.402±0.011**	**0.465±0.008**	**0.256±0.009**	**0.0755±0.0016**

The first six models are general-purpose models, and the last six are specifically designed for polymers. polyBERT, TransPolymer, and PerioGT are self-supervised learning models. The mean and standard deviation of the test root mean square error on three independent runs are reported. The best performance for each task is shown in bold.

Table 2 | Test the performance of different models on four downstream tasks, each supplemented with additional information

Dataset	Opv	Mar1	Mar2	Pe2
Number of polymers	**1203**	**520**	**520**	**271**
RF[43]	1.820±0.023	**0.209±0.002**	0.242±0.001	–
polyBERT	–	0.220±0.010	0.239±0.006	–
TransPolymer	2.056±0.018	0.277±0.024	0.276±0.012	0.638±0.021
PerioGT	**1.800±0.013**	**0.209±0.003**	**0.229±0.001**	**0.533±0.024**

Note that polyBERT does not support downstream tasks known as Opv and Pe2, which require additional information. The mean and standard deviation of the test root mean square error on three independent runs are reported. The best performance for each task is shown in bold.

information on the datasets, please refer to Supplementary Table 1-4.

Table 1 and Supplementary Table 12 summarize the test performance of PerioGT and baseline models on the eight regression datasets. Observations from Table 1 are as follows: (1) PerioGT outperforms the baseline supervised and self-supervised models on all datasets. (2) Standard descriptors and graph neural networks applicable to small molecules generally perform worse than methods specifically designed for polymers. For example, although MPNN and Antoniku's method (adding an edge between the head and tail of RU graph) share the same backbone (MPNN), Antoniku's method performs better on most tasks, indicating the heterogeneity of polymers and the

need to design methods specifically for polymers.

Table 2 demonstrates the test performance of PerioGT and baselines on more complex tasks. In these tasks, each sample either includes additional information along with a chemical structure or consists of a copolymer. Observations from Table 2 indicate that PerioGT outperforms baselines across all four datasets. The results suggest that PolyGraph endows PerioGT with a flexible mechanism to effectively integrate additional information into the pre-trained model, enhancing its adaptability to complex tasks involving additional information.

Overall, the results suggest that PA-based contrastive learning and PGFT enhance the model's generalization performance across diverse downstream tasks of polymers. Therefore, the notable generalization capability of PerioGT shows its potential for application in high-throughput screening of polymers.

PerioGT achieves a better understanding of the periodicity of polymers

After verifying PerioGT's excellent performance on multiple downstream tasks, we evaluated the model performance in capturing periodic features, coping with complex structural changes, and maintaining robust representation. In the following section, we will conduct our analysis from two perspectives: PA-level similarity analysis and instance-level similarity analysis.

PA-level similarity analysis. During the training process, a PA-based contrastive loss was applied to introduce the periodicity prior. To test the model's generalization capability for polymers that were not encountered in pre-training, we performed PA for all samples in the Tg dataset five times, and the average similarity between original RUs and augmentations was calculated to obtain the similarity matrix. As shown in Fig. 2a, the similarity matrix is visualized using block aggregation. Additionally, we analyzed the similarity distribution between the original RUs and their augmentations across the entire dataset, as illustrated in Fig. 2b. It is evident that PerioGT accurately identifies different augmentations of the same polymer, with an average similarity of 0.84, significantly better than the baseline methods. Further, we subsequently calculated the Davies-Bouldin index (DB index) of the representations above. From the observations in Fig. 2c, PerioGT demonstrates superior clustering performance, whereas baseline methods are more sensitive to PA. On this basis, we designed a new task in which RUs sampled from the same polymer were treated as a single class, while RUs from different polymers were treated as distinct classes. The goal of this task was to train a model for multiclass classification using the generated representations. To better highlight the effectiveness of representations, we employed a simple K-nearest neighbors (KNN) classifier. As shown in Fig. 2d, PerioGT achieved an accuracy exceeding 80%, while the accuracy of the other two SSL methods remained below 40%. Additionally, we investigated the generalization ability of PerioGT with varying numbers of MRUs. As shown in Fig. 2e, the similarity of all types of representations decreases to varying degrees as the number of MRUs increases. However, PerioGT consistently maintains a relatively high similarity, even when the maximum MRUs used in pre-training is limited to 3.

The above results suggest that PA-based contrastive learning can effectively handle varying arrangements and lengths of polymers. This property ensures that the model can still accurately recognize and predict polymers even when slight variations occur in their form.

Instance-level similarity analysis. Subsequently, we presented four pairs of instances in the Tg dataset to visually show the effect of PA-based contrastive learning. We used t-SNE algorithm to

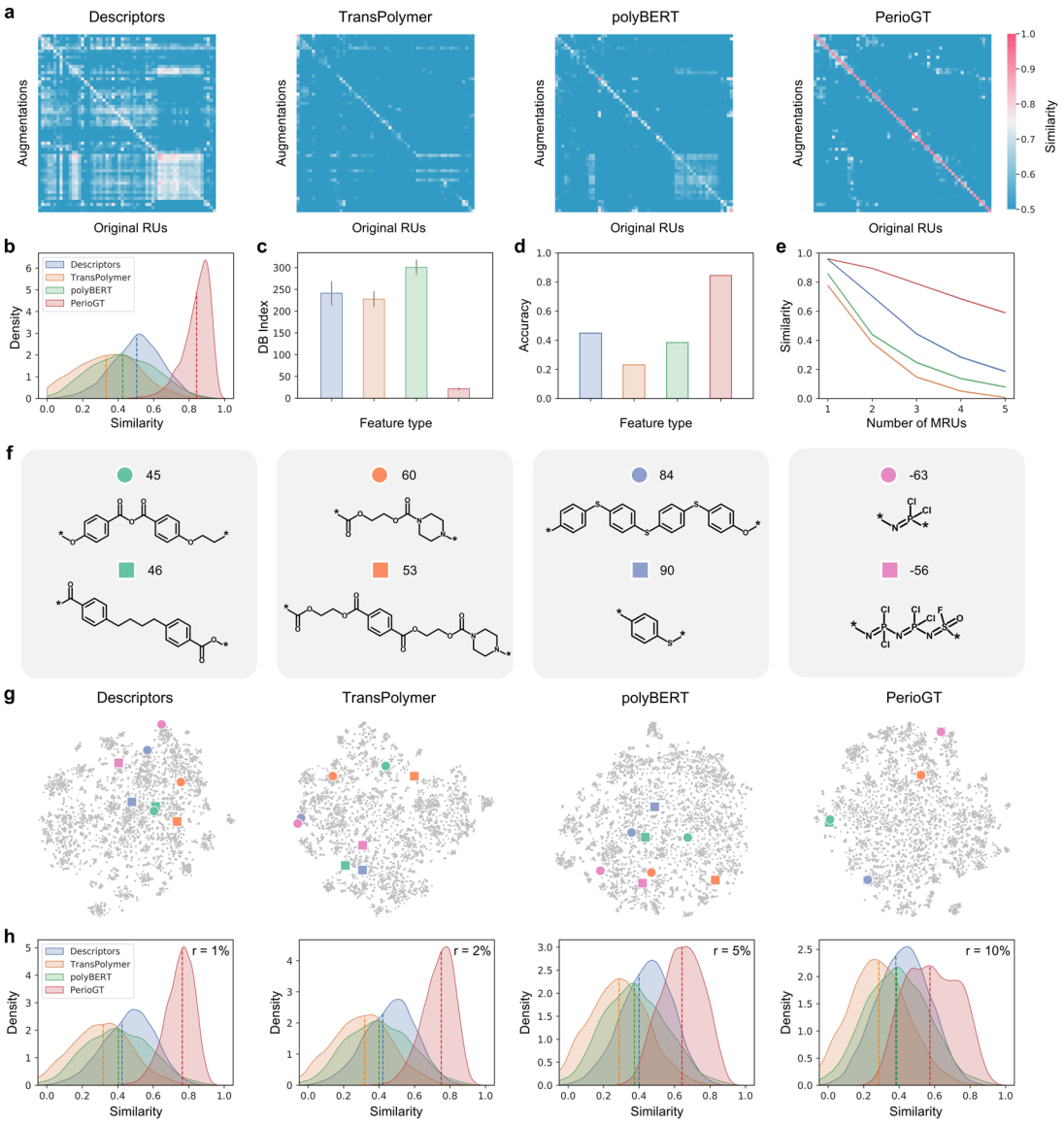

Fig. 2 | Similarity analysis. a, Similarity matrix of representations generated by different models. Each polymer in Tg dataset undergoes five random augmentations via PA, and their representations are extracted using the pretrained models. The similarity matrix is visualized by block aggregation. **b**, Similarity distribution of the generated representations. **c**, Alignment analysis. T-SNE algorithm is used to reduce the dimensionality of representations, and DB index is used to quantify their clustering performance. A lower DB index indicates better separation performance. **d**, KNN classification accuracy. One sample from each class is selected as the test set. **e**, The variation of similarity with the number of MRUs. **f**, Four groups of structurally similar polymers. Each sample is labeled with a unique color or symbol, and the corresponding glass transition temperature is noted to the right of the symbol. **g**, t-SNE visualization of polymer representations obtained from different methods. The marked points represent the corresponding samples in f. **h**, Similarity distribution of different replacement rate.

map the polymer representations learned by the pre-trained model into a two-dimensional space in Fig. 2d-e.

The polymers in each pair have similar structures. In the first pair, the structures of the two polyanhydrides are extremely similar: the former contains two ether bonds within the alkyl chain

between the two benzene rings, while the latter has only an alkyl chain between the rings, with all other structural features being identical. They have similar polarity and rigidity, similar mobility of chain segments, and thus closely matched glass transition temperatures (45°C and 46°C respectively). Despite their variability of MRU due to differences in the order of structure arrangement, PerioGT still successfully captures similar representations of these polymers. In contrast, baseline methods encode them into distant positions in the latent space. In the later pairs, the two polymers exhibit a high similarity in backbone and thus a high similarity in glass transition temperatures, so PerioGT extracts similar representations. However, due to the difference of MRU size, the representations generated by baseline methods are quite different. The results above indicate that, in addition to identifying variants of the same polymer, PA-based contrastive learning forces the model to effectively learn the decisive features of polymers from their structure, without being disturbed by minor changes such as the size of MRUs or minor atomic substitutions. In summary, PerioGT has a better understanding of the periodicity of polymers.

To show this statistically, we applied PA to all samples in the dataset followed by random replacement of some atoms. Although these modified polymers do not necessarily correspond to samples in the Tg dataset, we consider these augmentations analogous to the four instance groups discussed earlier. The similarity distribution between the original and augmented RUs was then calculated as shown in Fig. 2h. The results indicate that, as expected, similarity decreases with an increasing replacement rate. However, PerioGT consistently maintains the highest similarity values, even after minor perturbations, compared to the baseline models.

A well-organized representation space of PerioGT improves predictions

After the validation of model's understanding of polymer periodicity, we further analyzed its representation space. Recent research in representation learning has demonstrated that a fair representation space should filter out unnecessary details while retaining as much information as possible, which involves two aspects: alignment and uniformity[44]. Alignment implies that the model encodes similar samples into similar representations. Here we categorize it into attribute alignment and structural alignment to distinguish between the relationship of representation and polymer properties and the relationship of representation and polymer structure. Uniformity indicates that features are evenly distributed across the representation space, thereby maximizing information retention.

Attribute alignment. Attribute alignment tends to encode polymers with similar properties into similar representations, which verifies whether the pre-training task can effectively guide the model to capture highly expressive features. Thus, we calculated the Mutual Information (MI) between representations and several fundamental physicochemical properties of polymers, as illustrated in Fig. 3a. The results indicate that PerioGT exhibits the highest correlation between principal component and properties compared to existing SSL models (TransPolymer and polyBERT) and standard descriptors.

To further analyze attribute alignment, the k-Nearest Neighbors (KNN) algorithm was employed. Specifically, 10% of the samples from each dataset were randomly selected as test data, while the remaining 90% were used to train a KNN regressor. The trained KNN model was then evaluated on the reserved test samples to assess its performance. As shown in Fig. 3b, KNN regressor trained on representations generated by PerioGT outperforms baseline models on most attributes, indicating that close representations have similar properties.

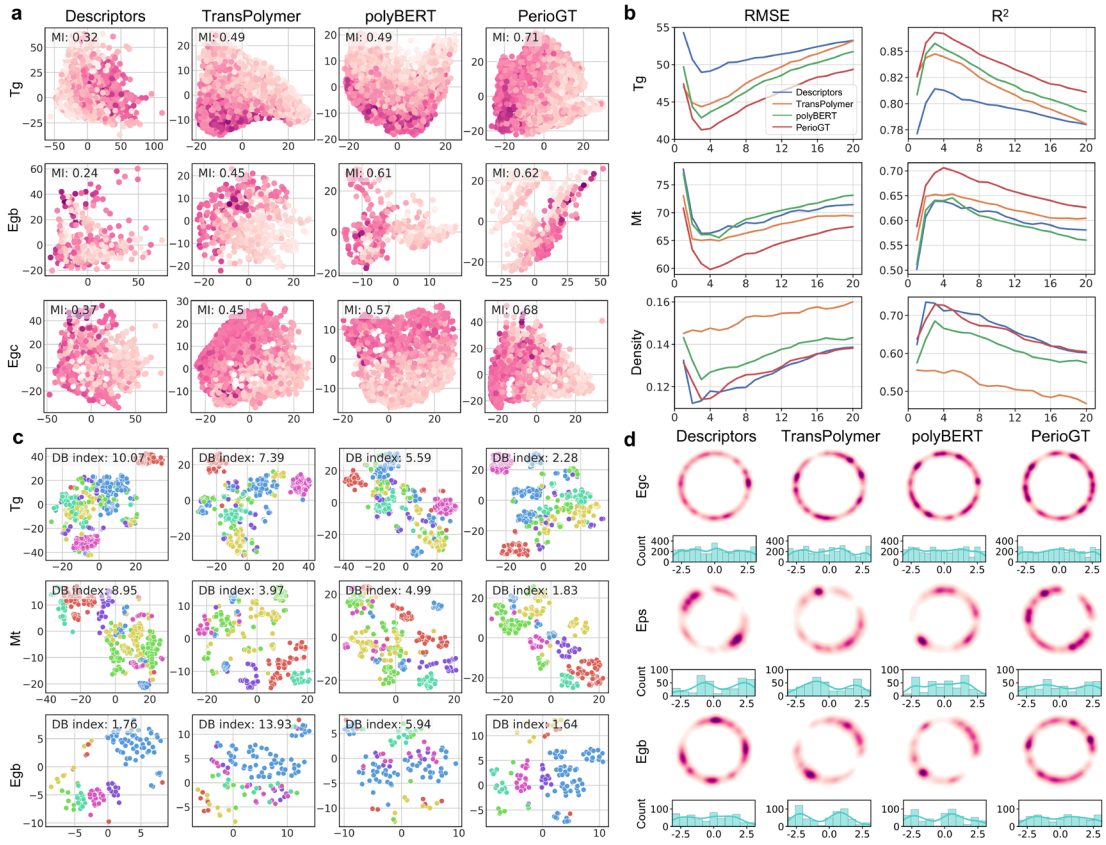

Fig. 3 | Attribute alignment, structural alignment and uniformity analysis. a, Mutual information between principal components and attributes. Points are colored with the value of the attribute. **b**, Performance of KNN regression using the generated representations. The mean values at different k values on three independent runs are shown. **c**, t-SNE visualization of eight major molecular scaffolds across different tasks. Points are colored based on scaffold type. A lower DB index indicates better separation performance. **d**, Uniformity analysis. We plot polymer representations with Gaussian kernel density estimation (KDE) in $\mathbb{R}^2$ and a histogram of radians (i.e., arctan2(y, x) for each data $(x, y) \in S^1$).

The above results demonstrate that PerioGT effectively extracts rich physicochemical knowledge of polymers after applying the proposed pre-training strategy. Moreover, the well attribute alignment in multiple tasks implies that the model possesses strong generalization capabilities, enabling it to quickly adapt to unseen downstream tasks through simple fine-tuning.

Structural alignment. Structural alignment favors models that encode samples with similar chemical structures into adjacent positions in the latent space. To characterize the structural alignment of PerioGT, we followed the method employed by ref [37]. We extracted the Murcko Scaffold of each sample from downstream tasks, which retains only the core structure (ring systems and linkers), disregarding substituent groups. Polymers sharing the same chemical scaffold are likely to exhibit similar properties. We then visualized eight major scaffolds using t-SNE and calculated the DB index to quantify the structural alignment. As illustrated in Fig. 3c, PerioGT consistently exhibits a lower DB index, indicating superior clustering performance. This observation suggests that the representations generated by PerioGT are invariant to unnecessary details.

Uniformity. Recent studies in representation learning have revealed that representations should retain as much information as possible. In other words, the pre-trained encoder should encode polymers uniformly into the latent space. To this end, we qualitatively explored the uniformity of PerioGT using methods proposed in previous studies[37,44]. First, representations were generated by the pre-trained model for specific downstream tasks. Next, we reduced the dimensionality of these generated representations with t-SNE and performed L2 normalization. We visualized the normalized representation through Gaussian Kernel Density Estimation (KDE) and the probability density of radians on the unit circle. As shown in Fig. 3d, the features generated by PerioGT are more uniformly distributed on the unit circle, indicating its superior uniformity.

Ablation study

In the aforementioned experiments, PerioGT demonstrates state-of-the-art performance, primarily due to the incorporation of several advanced techniques, including PA-based contrastive learning, PolyGraph (introduction of virtual nodes), masked modeling, and periodicity prompt. To investigate the individual contributions of these techniques to the framework, we conducted ablation studies. Consequently, we constructed six models: PerioGT-w/o VN, PerioGT-w/o MNM, PerioGT-w/o CL, PerioGT-w/o PA, PerioGT-linear probe and PerioGT-prompt tuning. We evaluated the performance of these six models in eight downstream tasks. To maintain consistency with previous representation learning studies and ensure fair comparisons, we froze the weights of the backbone and trained a projection head only.

The results (Supplementary Table 13) indicate that each component of PerioGT contributes to performance in most tasks. Notably, PerioGT consistently outperforms PerioGT-w/o CL and PerioGT-w/o PA in test performance, highlighting the importance of incorporating periodicity priors into the model.

Case study: the potential of PerioGT is exploited in antimicrobial polymer discovery

Finally, to validate the real-world applicability of PerioGT, we applied it to the discovery of antimicrobial polymers (AMPs) in web-lab experiments. The emergence of antibiotic-resistant bacteria, driven by antibiotic misuse, poses a significant threat to human health. Thus, the discovery of new compounds with antimicrobial activity is increasingly prioritized[45,46]. Antimicrobial peptides are considered to be antimicrobial compounds that do not easily induce resistance. However, they are limited by their high synthetic cost and low in vivo stability[47,48]. To this end, many researchers introduce cationic and hydrophobic groups into synthetic polymers to mimic the physicochemical characteristics of antimicrobial peptides[49-54]. Most importantly, AMPs are normally synthesized from inexpensive raw materials through a cost-effective, scalable chemical synthesis process, making them promising alternatives to antimicrobial peptides.

However, unlike the abundant data accumulated for antimicrobial peptides, there is currently no data available at scale for AMPs. Therefore, we constructed a combinatorial polymer library using Michael addition[55], as shown in Fig. 4a. We then tested the antimicrobial properties of the samples through high-throughput synthesis and characterization techniques. More details about the library construction and wet-lab experiments can be found in the Methods section.

Following this, PerioGT was fine-tuned with the constructed AMPs dataset. As demonstrated in Table 3, PerioGT achieves the best performance, which is consistent with the above findings. To predict AMPs with novel structures, we selected commercially available monomers that are

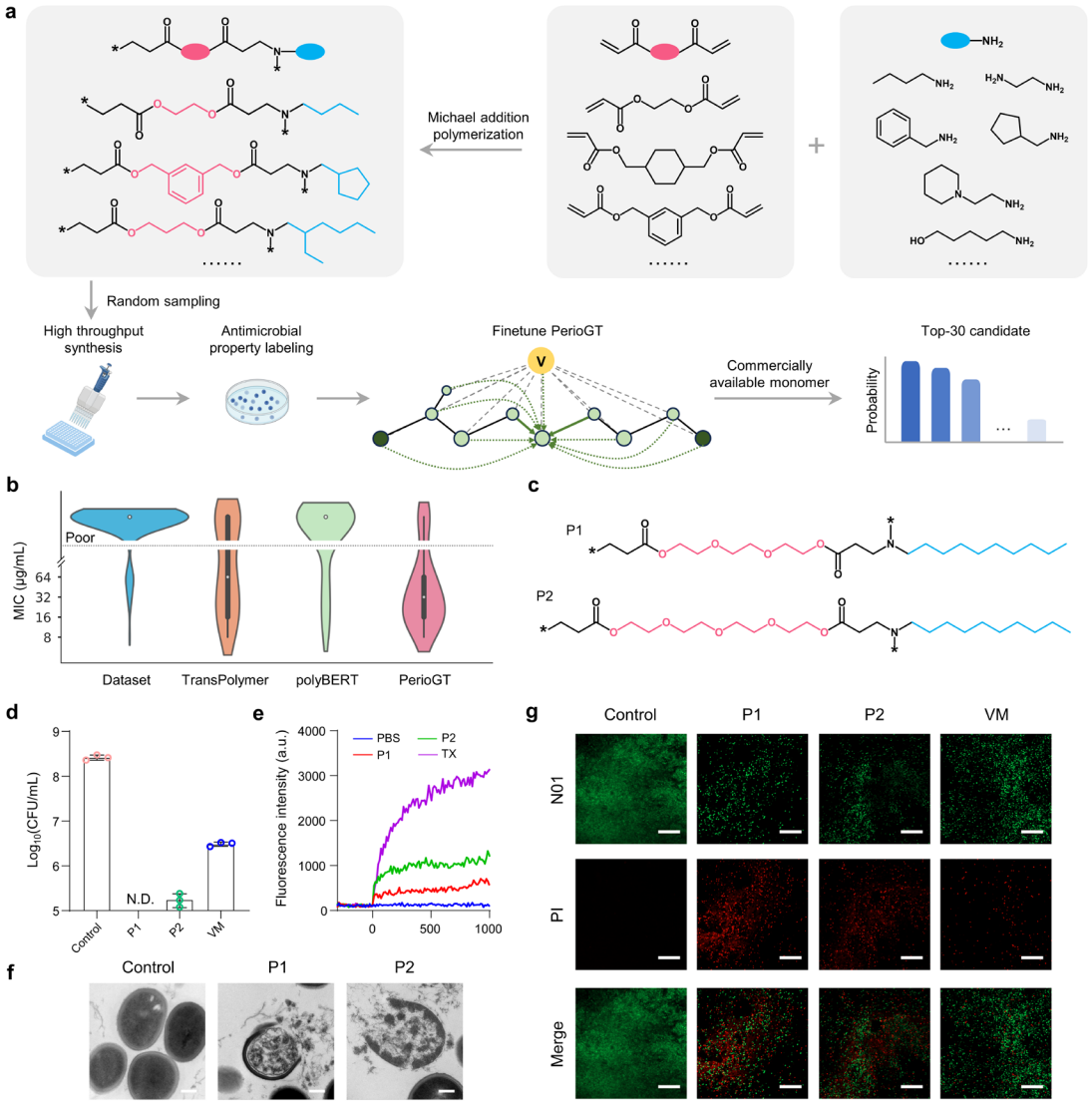

Fig. 4 | Application of PerioGT in AMP discovery. a, Screening workflow. The library to be screened was constructed based on combinatorial chemistry and Michael addition polymerization. Any pair-wise combination between diacrylates (pink) and amines (blue) can generate a unique polymer. We randomly sampled 150 polymers for wet-lab synthesis and antimicrobial activity labeling. PerioGT was then fine-tuned on the constructed dataset. We employed the fine-tuned PerioGT to predict unlabeled Michael addition products whose monomers were commercially available, and the top-30 candidate with the highest positive probability were selected for further characterization. **b**, Distribution of MIC in the training set and top-30 predictions by different self-supervised models. The contour shows the kernel density estimation, with a white dot for the median, a thick bar for the interquartile range, and thin lines for the 95% confidence intervals. A lower MIC indicates a better antimicrobial property. The part above the dashed line indicates no antimicrobial activity (MIC≥128 μg/mL). **c**, The two AMPs with the lowest MIC (8 μg/mL) predicted by PerioGT and evaluated by wet-lab experiments. **d**, MRSA colony counting after incubation with P1-P3 for 6 h. **e**, Membrane potential disturbance induced by the polymers. **f**, TEM characterization of MRSA after incubation with P1-P3 for 5 min. Scale bar: 200 nm. g, Live/Dead staining test results. Scale bar: 20 μm.

Table 3 | Test performance of different models on the constructed antimicrobial polymers dataset

Model	ROC-AUC	PR-AUC
RF	0.734±0.145	0.290±0.180
MPNN	0.395±0.075	0.214±0.027
GAT	0.721±0.214	0.214±0.145
Graphormer	0.659±0.150	0.143±0.076
LiGhT	0.630±0.186	0.192±0.163
GraphGPS	0.721±0.170	0.250±0.160
Arora et al.	0.871±0.124	0.397±0.210
Antoniuk et al.	0.519±0.174	0.147±0.147
Aldeghi et al.	0.699±0.175	0.255±0.230
TransPolymer	0.768±0.196	0.400±0.260
polyBERT	0.708±0.240	0.240±0.198
PerioGT	**0.873±0.144**	**0.479±0.295**

The mean and standard deviation of the root-mean-square error on ten different splits are reported.

reasonably priced and of high purity. Polymers are then virtually generated by simulating the anticipated Michael addition reactions between the selected amines and diacrylates, used to construct a library of polymers for screening. Next, the fine-tuned model was used to predict the antimicrobial activity of the entire polymer library. Based on the predictions, we selected the top 30 polymers with the highest predicted positive probabilities and evaluated them through wet-lab experiments to verify the model's reliability (Supplementary Table 11). We also compared the performance of PerioGT with two published SSL models. As shown in Fig. 4b, with a minimum inhibitory concentration (MIC) threshold of 64 μg/mL, PerioGT achieves a screening success rate of 83% (25 out of 30), much higher than TransPolymer at 57% (17 out of 30) and polyBERT at 20% (6 out of 30). These results demonstrate that PerioGT exhibits superior accuracy and efficiency in identifying AMPs.

Through the validation of PerioGT, two polymers were identified with MIC at 8 μg/mL (labeled as P1 and P2 in Fig.4c). The antibacterial activity of P1-P2 was further investigated by incubation with MRSA for 9 h. As shown in Fig. 4d, P1 is capable of eliminating MRSA by over 6 orders of magnitude, while P2 reduces the MRSA colony count by 3 orders of magnitude. In comparison, vancomycin hydrochloride (VM), a widely applied clinical antibiotic especially against gram-positive bacteria such as MRSA, eradicates the MRSA colony count by 2 orders of magnitude. MRSA cells treated with the lead polymers were then subjected to Live/dead staining assays. Propidium iodide (PI) is capable of penetrating impaired bacterial membranes and binding to DNA to emit red fluorescence, while N01 probes stain bacteria regardless of their state. Both P1-2 and VM-treated samples exhibit pronounced red fluorescence, while the control group is only stained by N01 probes (Fig. 4g). The antibacterial mechanism of P1-P2 was further explored. Membrane depolarization probe DiSC3(5) concentrates in bacterial membranes and quenches due to the high concentration. It would emit fluorescence upon disturbance of membrane potential. Triton X-100 (TX), a potent surfactant, along with P1-P2, induces evident membrane potential disturbances in MRSA cells (Fig. 4e). However, VM fails to disrupt membrane potential balance, as its antibacterial

mechanism is known as the inhibition of bacterial cell wall synthesis. TEM observations also reveal severe membrane dissociation and the release of cytoplasmic contents from MRSA cells (Fig. 4f). These findings underscore the strong interaction between P1-P2 and the bacterial cell membrane.

Conclusion

Deep learning modeling for polymers is of great importance, but due to their complex structures, developing effective strategies is still challenging. In this study, we propose a periodicity-aware deep learning method for polymer representation learning and property prediction. In PerioGT, a novel contrastive learning pre-training strategy based on Periodicity Augmentation (PA) is introduced as a periodicity prior. To emphasize the periodic structure of polymers, a periodicity prompt is incorporated into the graph in fine-tuning. Additionally, given that the properties of polymers are influenced by multilevel factors, a modular graph construction method named PolyGraph is proposed to introduce additional conditions in the form of virtual nodes. In comparison with baseline methods, PerioGT demonstrates superior performance in all downstream tasks. We further conclude that PerioGT has a better feature space quality and a better understanding of the periodic structure of polymers through attribute alignment, structural alignment, uniformity, and similarity analysis.

Notably, the effectiveness of PerioGT has also been validated in real-world applications. We applied PerioGT to the screening of antimicrobial polymers (AMPs). We constructed the training dataset through Michael-addition reaction and high-throughput experimental techniques. PerioGT was then used to predict polymers with novel structures, discovering 25 novel AMPs, with a success rate much higher than baselines. Additionally, two polymers with potent antibacterial properties were identified by PerioGT, and their antibacterial performance against MRSA was evaluated.

In summary, PerioGT provides an innovative approach to the periodic modeling of polymers and exhibits state-of-the-art performance in representation learning and polymer property prediction.

Methods

PolyGraph construction

In this work, we propose a polymer representation method named PolyGraph. Given a polymer fragment, it can be represented as $G=(V,E)$, where $V=\{v_i\}_{i\in[1,N_v]}$ and $E=\{e_{i,j}\}_{i,j\in[1,N_v]}$ are nodes (atoms) and edges (bonds), respectively. G is further transformed into a line graph $\hat{G}=(\hat{V},\hat{E})$ as the following steps. (1) Node creation: For each $e_{i,j}\in E$, create a corresponding node $\hat{v}_{i,j}\in\hat{V}$. (2) Edge creation: Two nodes in $\hat{V}$ are connected by an edge in $\hat{E}$ if and only if the corresponding two edges in G share a common vertex. The features of $\hat{v}_{i,j}$ are initialized to $\boldsymbol{h}_{v_{i,j}}\in\mathbb{R}^d$ which is formulated as:

$$\boldsymbol{h}_{v_{i,j}}=\text{concat}(\boldsymbol{W}_v\boldsymbol{x}_i+\boldsymbol{W}_v\boldsymbol{x}_j,\boldsymbol{W}_e\boldsymbol{x}_{i,j}), \tag{1}$$

where is d the embedding dimension, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ denote atom features of v_i and v_j respectively, $\boldsymbol{x}_{i,j}$ is the bond feature of $e_{i,j}$, $\boldsymbol{W}_v\in\mathbb{R}^{0.5d\times d_{atom}}$ and $\boldsymbol{W}_e\in\mathbb{R}^{0.5d\times d_{bond}}$ are trainable weights. Line graphs emphasize the importance of chemical bonds in polymers. Recently, line graph-based approaches have shown promising performance on various chemistry-related tasks[41,56]. Thus, we adopted line graphs in the PerioGT implementation.

N_{VN} virtual nodes are then added to $\hat{V}$. The features of virtual nodes $\hat{v}_{\text{expt}}$ and $\hat{v}_{add}$ are initialized with expert knowledge of polymers $\boldsymbol{x}_{\text{expt}}$ and additional information $\boldsymbol{x}_{add}$, respectively, to provide

supplementary details on polymerization degree, molecular weight distribution, crosslinking degree, end groups, and any available conditions. Notably, $\boldsymbol{x}_{add}$ is only introduced in fine-tuning due to its weak correlation with the chemical structure of MRUs. Considering the time and space complexity of polymer expert knowledge computation, we here approximate the physicochemical properties of the multimer (n=3) calculated by Mordred as $\boldsymbol{x}_{\text{expt}}$ [57].

Self-supervised pre-training

In this work, we employ a multi-scale SSL strategy to learn features at different levels of polymers. The dataset used in pre-training can be found in Supplementary Information. PA-based contrastive learning is proposed to incorporate a periodicity prior. PA-based contrastive learning aims to differentiate between augmentations of the same polymer and those of different polymers. For each polymer, we randomly and independently sample two fragments via PA and convert them into PolyGraphs $\hat{G}_i, \hat{G}_i'$ as a positive pair, while the fragments in the same minibatch sampled from other polymers are treated as negative pairs. To prevent the model from over-relying on obvious differences and to promote the learning of essential features, we introduce perturbations to PolyGraph. For physical nodes, we randomly mask 50% of the nodes in both positive and negative samples. The masking method follows the original BERT approach: for each node to be masked, there is an 80% chance of replacing it with a special node, a 10% chance of replacing it with a random node, and a 10% chance of leaving it unchanged[58]. For virtual nodes, we similarly mask 50% of the dimensions randomly. For numerical features, we replace values with Gaussian noise. For categorical features (e.g., binary values 0 or 1), we replace the feature with its opposite.

The graph encoder $f(\cdot)$ with the readout operator is used to extract graph representations followed by a non-linear projector $g_{\text{PACL}}(\cdot)$ to map representations into a latent space, generating representation vectors z_i, z_i'. Info NCE loss is applied to distinguishes the positive pair (z_i, z_i') from negative pairs $\bigcup\{\{(z_i, z_k),(z_i, z_k')\} \mid k \neq i\}$ for each PolyGraph $\hat{G}_i$, which is formulated as follow[59,60]:

$$\mathcal{L}_{\text{InfoNCE}}(i) = -\log \frac{\exp(z_i^{\top} z_i' / \gamma)}{\sum_{k \neq i} \exp(z_i^{\top} z_k' / \gamma) + \sum_k \exp(z_i^{\top} z_k' / \gamma)}, \tag{2}$$

where γ is a temperature hyperparameter. $\mathcal{L}_{PACL}$ is the averaged loss for all n samples in a mini batch:

$$\mathcal{L}_{\text{PACL}} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{\text{InfoNCE}}(i) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}'_{\text{InfoNCE}}(i), \tag{3}$$

where $\mathcal{L}'_{\text{InfoNCE}}(i)$ denotes the symmetric loss $\mathcal{L}_{\text{InfoNCE}}(i)$ for the paired positive sample $\hat{G}_i'$.

We hypothesize that PA-based contrastive learning enables the model to better understand the long-range structure of polymers, while Masked Node Modelling (MNM) helps the model learn the local features of polymers. Therefore, another non-linear projector $g_{\text{MNM}}(\cdot)$ maps the perturbed physical nodes into a latent space for MNM. Meanwhile, we here also reconstruct the virtual nodes features $\hat{v}_{\text{expt}}$ based on the perturbed $\hat{G}_i$. The loss function for MNM is defined as[58]:

$$\mathcal{L}_{\text{MNM}} = \frac{1}{3}(\mathcal{L}_{\text{PN}} + \mathcal{L}_{\text{VN\_CLS}} + \mathcal{L}_{\text{VN\_REG}}), \tag{4}$$

where $\mathcal{L}_{\text{PN}}$ represents the reconstruction loss of physical nodes (CrossEntropyLoss), $\mathcal{L}_{\text{VN\_CLS}}$ represents the reconstruction loss of categorical features in virtual nodes (BCELoss), and $\mathcal{L}_{\text{VN\_REG}}$

represents the reconstruction loss of numerical features in virtual nodes (MSELoss).

The overall training objective combines the PA-based contrastive loss, the MNM loss, and the reconstruction loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{PACL}} + \mathcal{L}_{\text{MNM}}. \tag{5}$$

There are three versions of PerioGT implemented: PerioGT-base (100M), PerioGT-small (30M), and PerioGT-large (250M), as described in Supplementary Table 6 and 7.

Graph encoder architecture

Due to the large chemical structures of polymers, we employ Graph Transformer (GT) as the graph encoder. GT can effectively capture long-range dependencies in polymers through attention mechanism and mitigate the over-smoothing problem[40].

Unlike vanilla Transformer, GT needs to incorporate graph structure information to enhance the graph structure awareness of the Transformer model. We implement LiGhT[41] and GraphGPS[42] as the graph encoder $f(\cdot)$ here. The structural encoding of LiGhT includes Path Encoding (PE) and Distance Encoding (DE). PE is used to encode shortest path between nodes to represent their structural relationships within the graph. And DE encodes the direct pairwise distances between nodes. Formally, the path feature $a_{i,j}^{p}$ between $\hat{v}_i$ and $\hat{v}_j$ in the path attention matrix $\boldsymbol{A}^{p} \in \mathbb{R}^{N_{\hat{v}} \times N_{\hat{v}}}$ is defined as follow:

$$a_{i,j}^{p} = \boldsymbol{W}^{p} \frac{1}{N_p} \sum_{n=1}^{N_p} \boldsymbol{W}_n^{p} \boldsymbol{h}_{v_n} \tag{6}$$

where $\boldsymbol{h}_{v_n}$ represents the feature of the n-th node in the shortest path between $\hat{v}_i$ and $\hat{v}_j$, N_p represents the number of nodes in the path, $\boldsymbol{W}_n^{p} \in \mathbb{R}^{d_p \times d}$ and $\boldsymbol{W}^{p} \in \mathbb{R}^{1 \times d_p}$ are trainable weights. Similarly, the distance feature $a_{i,j}^{d}$ between $\hat{v}_i$ and $\hat{v}_j$ in the distance attention matrix $\boldsymbol{A}^{d} \in \mathbb{R}^{N_{\hat{v}} \times N_{\hat{v}}}$ is calculated as follow:

$$a_{i,j}^{d} = \boldsymbol{W}_2^{d} \sigma(\boldsymbol{W}_1^{d} d_{i,j}) \tag{7}$$

where $d_{i,j}$ represents the distance between $\hat{v}_i$ and $\hat{v}_j$, $\boldsymbol{W}_1^{d} \in \mathbb{R}^{d_d \times 1}$ and $\boldsymbol{W}_2^{d} \in \mathbb{R}^{1 \times d_d}$ are trainable weights. GT follows a similar multi-head attention in vanilla Transformer. Specifically, for a given feature matrix $\boldsymbol{H}^{(l)} \in \mathbb{R}^{N_{\hat{v}} \times d}$, the multi-head attention mechanism MHSA$(\cdot)$ is formulated as follow[61]:

$$\begin{aligned}
&\boldsymbol{Q}_k = \boldsymbol{H}^{(l)} \boldsymbol{W}_k^{Q}, \boldsymbol{K}_k = \boldsymbol{H}^{(l)} \boldsymbol{W}_k^{K}, \boldsymbol{V}_k = \boldsymbol{H}^{(l)} \boldsymbol{W}_k^{V}, \\
&\boldsymbol{A}_k = \text{softmax}\left(\frac{\boldsymbol{Q}_k \boldsymbol{K}_k^{\top}}{\sqrt{d_k}} + \boldsymbol{A}^{p} + \boldsymbol{A}^{d}\right), \boldsymbol{O}_k = \boldsymbol{A}_k \boldsymbol{V}_k, \\
&\boldsymbol{H}_{\text{attn}} = \text{LayerNorm}(\text{concat}(\boldsymbol{O}_1, \boldsymbol{O}_2, ..., \boldsymbol{O}_{N_h}) + \boldsymbol{H}^{(l)}),
\end{aligned} \tag{8}$$

where $\boldsymbol{W}_k^{Q}, \boldsymbol{W}_k^{K}, \boldsymbol{W}_k^{V} \in \mathbb{R}^{d \times d_k}$ are trainable weights of k-th head at layer l, $d_k = d / h$ is the dimension of each attention head, h stands for the number of heads. The multi-head attention mechanism computes queries, keys, and values using linear transformations, calculates attention scores, and then combines the weighted values across all heads with a residual connection to produce the intermediate matrix $\boldsymbol{X}_{\text{attn}}$. A fully connect network (FFN) is then applied at each sub-layer:

$$\boldsymbol{H}^{(l+1)} = \text{LayerNorm}(\sigma(\boldsymbol{H}_{\text{attn}}\boldsymbol{W}_1^{(l)})\boldsymbol{W}_2^{(l)} + \boldsymbol{H}_{\text{attn}}), \tag{9}$$

where $\boldsymbol{W}_1^{(l)} \in \mathbb{R}^{d\times 4d}$, $\boldsymbol{W}_2^{(l)} \in \mathbb{R}^{4d\times d}$ are trainable weights of FFN, $\sigma(\cdot)$ is the GELU activation function.

In the GraphGPS, a local graph neural network is employed to capture local structural features. For a given node feature vector $\boldsymbol{h}_v^{(l)}$, $\boldsymbol{h}_v^{(l+1)}$ is formulated as follow:

$$\begin{aligned} \boldsymbol{m}_v^{(l)} &= \text{aggregate}(\{\text{message}(\boldsymbol{h}_v^{(l)}, \boldsymbol{h}_u^{(l)}, \boldsymbol{e}_{uv}^{(l)}) : u \in \mathcal{N}(v)\}), \\ \boldsymbol{h}_v^{(l+1)} &= \text{update}(\boldsymbol{h}_v^{(l)}, \boldsymbol{m}_v^{(l)}), \end{aligned} \tag{10}$$

where $\text{message}(\cdot)$, $\text{aggregate}(\cdot)$, $\text{update}(\cdot)$ is the message function, aggregation function and update function in PNA[62], respectively. Then, a global self-attention mechanism is used to capture long-range dependencies:

$$\boldsymbol{H}^{(l+1)} = \text{MHSA}(\boldsymbol{H}_{\text{local}}^{(l+1)}), \tag{11}$$

where $\boldsymbol{H}_{\text{local}}^{(l+1)}$ is the node feature matrix, $\text{MHSA}(\cdot)$ denotes the multi-head attention.

Finally, a readout operator is applied to extract a graph-level representation:

$$\boldsymbol{h}_{\hat{G}} = \text{concat}(\text{mean}(\{\boldsymbol{h}_i^{\text{phys}}\}), \boldsymbol{h}^{\text{expt}}, \boldsymbol{h}^{\text{add}}), \tag{12}$$

where $\boldsymbol{h}_i^{\text{phys}}, \boldsymbol{h}^{\text{expt}}, \boldsymbol{h}^{\text{add}}$ denote the representations of real nodes, expert knowledge nodes, additional information nodes, respectively. As shown in Supplementary Table 5, we compare the performance of LiGhT and GraphGPS on multiple downstream tasks. LiGhT is adopted in the subsequent analysis due to better robustness.

Prompt generation

To comprehensively aggregate the equivalent motifs in different RUs of polymers, we introduce a periodic prompt based on PA in fine-tuning. For a given fragment $\hat{G}$, PA is used to obtain a set of augmentations $\{\hat{G}_i\}_{i\in[1,N_{\text{aug}}]}$. For any node v in $\hat{G}$, we match it to the corresponding motifs in $\{\hat{G}_i\}_{i\in[1,N_{\text{aug}}]}$ using RDKit package[63], resulting in $\{v_k\}_{k\in[1,N_{\text{eq}}]}$. Then, we utilize the pre-trained GT to extract the corresponding node representation $\boldsymbol{H}_v^{\text{eq}} \in \mathbb{R}^{N_{\text{eq}}\times d}$. Multi-head self-attention layers are further applied to obtain the periodic prompt $\boldsymbol{h}_v^{\text{prompt}}$. Finally, $\boldsymbol{h}_v^{\text{prompt}}$ is added to the initial feature of as a prompt:

$$\begin{aligned} \boldsymbol{H}_v^{\text{attn}} &= \text{MHSA}(\boldsymbol{H}_v^{\text{eq}}), \\ \boldsymbol{h}_v^{\text{prompt}} &= \boldsymbol{H}_v^{\text{attn}}[0], \\ \boldsymbol{h}_v &= \boldsymbol{h}_v^{\text{initial}} + \alpha \boldsymbol{h}_v^{\text{prompt}}, \end{aligned} \tag{13}$$

where $\boldsymbol{H}_v^{\text{attn}}[0]$ represents the [PROMPT] token embedding which interacts with other nodes, α is a learnable weight.

In-silico experiment settings

Since polymer fragments are originally represented as SMILES, we used RDKit package to convert them into graphs, as well as to calculate atomic and bond features, as detailed in the Supplementary

Table 3 and 4. In SSL stage, Adam optimizer is utilized with an initial learning rate of 2×10^{-4} and a weight decay of 10^{-6}. The PolynomialDecayLR scheduler is employed to adjust the learning rate dynamically. The model is pre-trained with batch size of 288 for 100000 steps. In the fine-tuning of PerioGT, the pre-trained graph encoder weights are loaded, while projectors for pre-training are removed, and a randomly initialized task-related predictor is added for predicting specific properties. Grid search and early stopping is conducted for hyperparameter selection, and the model with the lowest RMSE on the validation set is selected. The specific search ranges of hyperparameters are provided in the Supplementary Information. The above experiments were implemented using Pytorch and DGL package on a Linux server with NVIDIA RTX4090 graphics processing units.

Wet-lab experiment details

Dataset construction. 12 diacrylates and 52 amines are selected to create a comprehensive library comprising 624 polymers (Supplementary Fig. 2 and Supplementary Table 10). The dataset is randomly sampled from the candidate pool to yield 150 combinations.

Polymer synthesis. The feeding ratio of diacrylates to amines is 1.2:1. Specifically, 40 μL of amine (3M in DMF) is pipetted into 48 μL of diacrylate (3M in DMF) in a vial. The reaction vial is heated to 90°C, and the reaction proceeded for the next 24 h. In the case of diacrylates with Boc-protecting groups, 80 μL of hydrofluoric acid is added to the solution after the completion of Michael addtion, and the vial is stirred overnight to remove protections. Once the reaction is completed, the raw product solution is purified by precipitating in diethyl ether.

Antimicrobial activity labelling of polymers. Purified polymers are redissolved at 2.56 mg/mL in DMSO. MRSA is cultured in Tryptic Soy Broth (TSB) at 37°C overnight. The bacterial suspension is then diluted and adjusted to a concentration at 10^5 Colony Forming Unit (CFU)·mL^{-1}. Subsequently, 195 μL of bacterial solution is added to 96-well plates, followed by the addition of 5 μL of polymer solution (final concentration at 64 μg/mL). The mixtures are allowed to incubate for 9 h at 37°C. Polymers capable of maintaining clear solutions are labeled as antibacterial combinations, while those resulting in turbidity are classified as non-antibacterial combinations. The biological characterization details can be found in supplementary information.

Reference

1 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).

2 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

3 Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, 6693 (2024).

4 Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).

5 Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023).

6 Zhang, H. *et al.* Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* **621**, 396-403 (2023).

7 Sumi, S., Hamada, M. & Saito, H. Deep generative design of RNA family sequences. *Nat. Methods* **21**, 435-443 (2024).
8 Stokes, J. M. *et al.* A deep learning approach to antibiotic discovery. *Cell* **180**, 688-702 (2020).
9 Wong, F. *et al.* Discovery of a structural class of antibiotics with explainable deep learning. *Nature* **626**, 177-185 (2023).
10 Wang, C. *et al.* Combinatorial discovery of antibacterials via a feature-fusion based machine learning workflow. *Chem. Sci.* **15**, 6044-6052 (2024).
11 Rinehart, N. l. A machine-learning tool to predict substrate-adaptive conditions for Pd-catalyzed C–N couplings. *Science* **381**, 965-972 (2023).
12 Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80-85 (2023).
13 Szymanski, N. J. *et al.* An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86-91 (2023).
14 Wang, Y., Wang, J., Cao, Z. & Barati Farimani, A. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.* **4**, 279-287 (2022).
15 Hu, W. *et al.* Strategies for pre-training graph neural networks. In *Proc. 8th International Conference on Learning Representations* (2020).
16 Zemin, L. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023).
17 Qiang, B. *et al.* Bridging the gap between chemical reaction pretraining and conditional molecule generation with a unified model. *Nat. Mach. Intell.* **5**, 1476-1485 (2023).
18 Yang, J., Tao, L., He, J. L., McCutcheon, J. R. & Li, Y. Machine learning enables interpretable discovery of innovative polymers for gas separation membranes. *Sci. Adv.* **8**, 13 (2022).
19 Gurnani, R. *et al.* AI-assisted discovery of high-temperature dielectrics for energy storage. *Nat. Commun.* **15**, 10 (2024).
20 Tamasi, M. J. *et al.* Machine Learning on a Robotic Platform for the Design of Polymer-Protein Hybrids. *Adv. Mater.* **34**, 12 (2022).
21 Jiang, S. L., Dieng, A. B. & Webb, M. A. Property-guided generation of complex polymer topologies using variational autoencoders. *NPJ Comput. Mater.* **10**, 13 (2024).
22 Arora, A. *et al.* Random Forest Predictor for Diblock Copolymer Phase Behavior. *ACS Macro Lett.* **10**, 1339-1345 (2021).
23 Kuenneth, C. & Ramprasad, R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nat. Commun.* **14**, 4099 (2023).
24 Xu, C., Wang, Y. & Barati Farimani, A. TransPolymer: a Transformer-based language model for polymer property predictions. *NPJ Comput. Mater.* **9**, 64 (2023).
25 Antoniuk, E. R., Li, P., Kailkhura, B. & Hiszpanski, A. M. Representing Polymers as Periodic Graphs with Learned Descriptors for Accurate Polymer Property Predictions. *J. Chem. Inf. Model.* **62**, 5435-5445 (2022).
26 Aldeghi, M. & Coley, C. W. A graph representation of molecular ensembles for polymer property prediction. *Chem. Sci.* **13**, 10486-10498 (2022).
27 Pham, T., Tran, T., Dam, H. & Venkatesh, S. Graph classification via deep learning with virtual nodes. Preprint at https://arxiv.org/abs/1708.04357 (2017).
28 Chen, C., Ye, W. K., Zuo, Y. X., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564-3572 (2019).
29 Lam, H. Y. I. *et al.* Application of variational graph encoders as an effective generalist algorithm

in computer-aided drug design. *Nat. Mach. Intell.* **5**, 754-764 (2023).

30 Duan, Y. *et al.* Enhancing molecular property prediction through task-oriented transfer learning: integrating universal structural insights and domain-specific knowledge. *J. Med. Chem.* **67**, 9575-9586 (2024).

31 Xu, D., Yen, I. E.-H., Zhao, J. & Xiao, Z. Rethinking network pruning-under the pre-train and fine-tune paradigm. In *Proc. 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2021).

32 Pengfei, L. *et al.* Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Computing Surveys* **55**, 1-35 (2021).

33 Li, X. L. S., Liang, P. & Assoc Computat, L. Prefix-tuning: optimizing continuous prompts for generation. In *Proc. 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (2021).

34 Menglin, J. *et al.* Visual prompt tuning. In *Proc. 17th European Conference on Computer Vision* (2022).

35 Taoran, F., Yunchao, Z., Yang, Y., Chunping, W. & Lei, C. Universal prompt tuning for graph neural networks. In *Advances in Neural Information Processing Systems 36* (2023).

36 Junhyun, L., Wooseong, Y. & Jaewoo, K. Subgraph-level universal prompt tuning. Preprint at https://arxiv.org/abs/2402.10380 (2024).

37 Fang, Y. *et al.* Knowledge graph-enhanced molecular contrastive learning with functional prompt. *Nat. Mach. Intell.* **5**, 542-553 (2023).

38 Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *Proc. 34th International Conference on Machine Learning* (2017).

39 Veličković, P. *et al.* Graph attention networks. In *Proc. 6th International Conference on Learning Representations* (2018).

40 Ying, C. *et al.* Do Transformers really perform bad for graph representation? In *Advances in Neural Information Processing Systems 34* (2021).

41 Li, H., Zhao, D. & Zeng, J. KPGT: knowledge-guided pre-training of graph Transformer for molecular property prediction. In *Proc. 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2022).

42 Rampášek, L. *et al.* Recipe for a general, powerful, scalable graph Transformer. In *Advances in Neural Information Processing Systems 35* (2022).

43 Patel, R. A., Borca, C. H. & Webb, M. A. Featurization strategies for polymer sequence or composition design by machine learning. *Mol. Syst. Des. Eng.* **7**, 661-676 (2022).

44 Wang, T. & Isola, P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *Proc. 37th International Conference on Machine Learning* (2020).

45 Murray, C. J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629-655 (2022).

46 Shang, Z., Chan, S. Y., Song, Q., Li, P. & Huang, W. The strategies of pathogen-oriented therapy on circumventing antimicrobial resistance. *Research* **2020**, 2016201 (2020).

47 Mookherjee, N., Anderson, M. A., Haagsman, H. P. & Davidson, D. J. Antimicrobial host defence peptides: functions and clinical potential. *Nat. Rev. Drug Discov.* **19**, 311-332 (2020).

48 Huang, J. *et al.* Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nat. Biomed. Eng.* **7**, 797-810 (2023).

49 Shabani, S. *et al.* Synthetic peptide branched polymers for antibacterial and biomedical applications. *Nat. Rev. Bioeng.* **2**, 343-361 (2024).

50 Zhou, M. *et al.* A dual-targeting antifungal is effective against multidrug-resistant human fungal pathogens. *Nat. Microbiol.* **9**, 1325-1339 (2024).

51 Pham, P., Oliver, S., Nguyen, D. T. & Boyer, C. Effect of Cationic Groups on the Selectivity of Ternary Antimicrobial Polymers. *Macromol. Rapid Commun.* **43**, 13 (2022).

52 Pham, P., Oliver, S., Wong, E. H. H. & Boyer, C. Effect of hydrophilic groups on the bioactivity of antimicrobial polymers. *Polym. Chem.* **12**, 5689-5703 (2021).

53 Phuong, P. T. *et al.* Effect of Hydrophobic Groups on Antimicrobial and Hemolytic Activity: Developing a Predictive Tool for Ternary Antimicrobial Polymers. *Biomacromolecules* **21**, 5241-5255 (2020).

54 Judzewitsch, P. R., Zhao, L., Wong, E. H. H. & Boyer, C. High-Throughput Synthesis of Antimicrobial Copolymers and Rapid Evaluation of Their Bioactivity. *Macromolecules* **52**, 3975-3986 (2019).

55 Furka, Á. Forty years of combinatorial technology. *Drug Discov. Today* **27** (2022).

56 Bai, P., Liu, X. & Lu, H. Geometry-aware line graph transformer pre-training for molecular property prediction. Preprint at https://www.arxiv.org/abs/2309.00483 (2023).

57 Moriwaki, H., Tian, Y.-S., Kawashita, N. & Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminformatics* **10**, 4 (2018).

58 Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional Transformers for language understanding. Preprint at https://www.arxiv.org/abs/1810.04805v2 (2018).

59 He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. Momentum contrast for unsupervised visual representation learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2020).

60 You, Y. *et al.* Graph contrastive learning with augmentations. Preprint at https://www.arxiv.org/abs/2010.13902v3 (2020).

61 Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems 30* (2017).

62 Corso, G., Cavalleri, L., Beaini, D., Liò, P. & c, P. V. Principal neighbourhood aggregation for graph nets. In *Advances in Neural Information Processing Systems 33* (2020).

63 Landrum, G. RDKit: Open-source cheminformatics. *Retrieved from http://www.rdkit.org* (2013).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2022YFB3807300), and the National Natural Science Foundation of China (52293381, 51933009).

Author contributions

Y.W. conceived the main idea and conducted the in-silico experiments. C.W. was responsible for the chemical synthesis and biological characterization. Y.W. and C.W. wrote the manuscript together. X.S. and T.Z. participated in the discussions and provided many suggestions for the wet-lab experiments. P.Z. and J.J. guided the whole project. All authors reviewed and approved the final manuscript.