Received: 26 June 2025 | Revised: 28 August 2025 | Accepted: 25 September 2025

DOI: 10.1002/inmd.70064

RESEARCH ARTICLE

Interdisciplinary MEDICINE Open Access

TargetGen-RNN model discovers novel antibiotic compound targeting drug-resistant *Staphylococcus aureus*

Shakeel Ahmad Khan[1] | **Adnan Shakoor**[2,3] | **Sadia Kanwal**[4]

[1]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[2]Department of Control & Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

[3]Center for Bio Systems and Machines, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

[4]Department of Biochemistry, University of Agriculture, Faisalabad, Pakistan

Correspondence
Shakeel Ahmad Khan and Adnan Shakoor.
Email: sakhan@polyu.edu.hk and adnan.shakoor0886@gmail.com

Funding information
Interdisciplinary Research Center for Biosystems and machines, Grant/Award Number: INBS2501

Abstract

The escalating threat of antibacterial resistance demands innovative approaches for the discovery of novel antibiotics. Identifying hit and lead compounds with unique scaffolds during the early phases of drug development remains a significant challenge. Although various generative models have been proposed to create drug-like molecules, their capacity to design wet-lab-validated target-specific compounds with novel scaffolds has been scarcely validated. Herein, we propose TargetGen-recurrent neural network (RNN), a state-of-the-art deep generative learning model designed to explore chemical space and generate novel tailor-made virtual compound libraries for specific biological targets. By leveraging a combination of transfer learning, temperature-modulated sampling, and stringent chemical validation, TargetGen-RNN was trained on 5.7 million drug-like compounds from the ZINC database and fine-tuned with 82 known *Staphylococcus aureus* DHFR inhibitors, yielding 28,708 structurally diverse and chemically viable novel, tailor-made molecules. Virtual screening, including QSAR analysis, pharmacophore mapping, molecular docking, molecular dynamic simulations, and multi-criteria decision analysis against the generated tailor-made compound library, led to the discovery of a potent antibiotic compound (SAK-2970) with a novel scaffold with high predicted antibiotic activity, binding affinity, and favorable ADMET profiles. SAK-2970 demonstrated remarkable in vitro bactericidal activity against *S. aureus*, strong biofilm inhibition and eradication capabilities, and exceptional efficacy against ciprofloxacin resistant *S. aureus*. In a mouse model of drug-resistant bacteremia, SAK-2970 significantly reduced bacterial load and improved survival rates with minimal systemic toxicity, underscoring its biocompatibility and therapeutic potential. These findings validate SAK-2970 as a promising candidate for developing antibiotic treatments targeting resistant bacterial infections and highlight TargetGen-RNN's powerful capability to generate hit compounds with novel scaffolds, advancing the frontier of antibiotic discovery.

KEYWORDS
antibacterial resistance, antibiotic discovery, deep generative models, novel molecular scaffolds, virtual screening

1 | INTRODUCTION

Antimicrobial resistance (AMR) poses an escalating threat to global health, with *Staphylococcus aureus*, particularly methicillin-resistant strains (MRSA), being at the forefront of this crisis. This versatile pathogen is responsible for a spectrum of diseases ranging from minor skin infections to life-threatening systemic conditions such as sepsis, endocarditis, and osteomyelitis.[1–6] Its remarkable adaptability and ability to acquire and disseminate resistance determinants have rendered it a formidable adversary in clinical settings.[7] The World Health Organization warns that if current trends persist, AMR could cause up to 10 million deaths annually by 2050, underscoring the urgent need for novel therapeutic strategies.[8] Historically, the treatment of *S. aureus* infections has relied on antibiotics, such as beta-lactams, glycopeptides (e.g., vancomycin), and lipopeptides[4–6] (e.g., daptomycin). However, the emergence of MRSA and vancomycin-intermediate/resistant *S. aureus* (VISA/VRSA) strains has significantly compromised their efficacy.[9,10] Compounding this challenge, *S. aureus*'s ability to form biofilms and engage in horizontal gene transfer exacerbates treatment difficulties, leading to increased morbidity, mortality, and healthcare costs.[11,12] The diminishing effectiveness of existing antibiotics highlights the imperative for new agents with novel mechanisms of action capable of circumventing traditional resistance pathways.[13]

Traditional antibiotic discovery methods, including high-throughput screening (HTS), structure-based drug design (SBDD), and fragment-based drug discovery (FBDD), have been instrumental but come with significant limitations.[14–16] HTS often yields low hit rates and demands substantial financial and time investments. Although SBDD and FBDD offer more focused approaches, they are constrained by the complexities of accurately modeling protein-ligand interactions in dynamic biological environments. Additionally, natural product screening faces challenges, such as the frequent rediscovery of known compounds and the synthetic complexity of natural products.[17] These obstacles necessitate the development of advanced methodologies to accelerate the discovery of effective antibiotics, particularly against multidrug-resistant (MDR) pathogens such as MRSA.

Recent advancements in generative deep learning (GDL) have revolutionized de novo molecular design by enabling the creation of novel compounds through data-driven approaches that bypass the traditional limitations.[18,19] Deep neural networks functioning as generative models eliminate the need for explicit design rules, thereby overcoming the challenges inherent in conventional techniques.[20] Various architectures have been explored, including variational autoencoders (VAEs),[21] generative adversarial networks (GANs),[22] graph convolutional networks (GCNs),[23] transformers,[24] and recurrent neural networks (RNNs),[25] which have demonstrated exceptional efficacy. By leveraging sequential molecular representations, such as SMILES strings, RNN-based models capture complex chemical syntax and semantics, facilitating the generation of valid and diverse molecular structures. However, many RNN-based GDL models are goal-directed, optimizing molecules to satisfy predefined objective functions.[26] This focus can lead to compounds that, despite scoring highly on computational metrics, may lack practical applicability owing to unrealistic physicochemical properties or synthetic infeasibility.[27,28] Furthermore, these models often lack experimental validation[19] and tend to generate molecules that are structurally similar to known active compounds, limiting the exploration of novel chemical spaces. Although conditional RNNs (cRNNs) enhance scaffold diversity through latent-space interpolation, they often lack precise target-specific optimization, restricting their ability to generate molecules that are closely aligned with specific biological targets.[29] The validation processes for these models primarily assess scaffold diversity, leaving room for improvement in ensuring the biological relevance of the generated compounds.

To surmount these limitations, we propose TargetGen-RNN, a GDL model based on a distribution-learning RNN that eliminates the need for explicit goal functions. TargetGen-RNN generates novel molecules by learning the chemical distribution of both large-source data and specific target data of active compounds. Our model leverages transfer learning,[30] early stopping, dropout regularization, optimized hyperparameters, and temperature-scaled sampling to effectively mitigate overfitting (a critical consideration given the limited size of fine-tuned data) and enhance the diversity and novelty of the generated compounds. We also integrated a comprehensive validation process utilizing RDKit for chemical plausibility and Tanimoto similarity to assess novelty and uniqueness, confirming that the generated molecules were valid and distinct from existing compounds. This integrative approach of transfer learning, controlled generative processes, and stringent validation empowers TargetGen-RNN to transcend prior constraints, there by providing a potent tool for generating novel, biologically relevant molecules while preserving the chemical diversity. We then applied TargetGen-RNN and de novo discovered antibacterial scaffolds targeting drug-resistant *S. aureus*, yielding SAK-2970, a novel molecule validated through HTVS, in vitro, and in murine bacteremia models. The success of TargetGen-RNN in generating structurally novel molecules with optimal pharmacokinetic profiles underscores its potential to overcome the limitations of traditional antibiotic discovery methods.

2 | MATERIALS AND METHODS

2.1 | Dataset preparation, TargetGen-RNN implementation, and de novo molecule generation

A total of 5.7 million drug-like compounds were retrieved from the ZINC database,[31] filtered by molecular weight (300–425 Da) and logP (2–3.5), and pre-processed using RDKit. Additionally, 82 known *S. aureus* DHFR inhibitors were collected from BindingDB.[32] DHFR was selected as the fine-tuning target owing to the enzyme's central role in *S. aureus* folate metabolism, involvement in trimethoprim resistance, and the availability of inhibitors with experimentally validated Ki values.[32–34] After standardization and deduplication, 5.6 million molecules were retained as the source data and 82 as the target data. TargetGen-RNN was implemented and trained to learn the structural features of the source data in the SMILES format. The model architecture comprised an embedding layer, an long short-term memory (LSTM) layer with 128 hidden units, and an output layer. The training was performed using the Adam optimizer for 30 epochs with early stopping. Fine-tuning was subsequently conducted on the target data to guide the generation of DHFR-specific inhibitors. The fine-tuned model generated novel SMILES. Chemical validity and uniqueness were ensured using the RDKit. The compounds were filtered based on scaffold novelty, drug-likeness (Lipinski, Veber, and Ghose rules), and physicochemical properties.[35–37] A final set of 6672 compounds underwent further computational screening (see Supplementary Sections S1.1–S1.5 for full details).

2.2 | QSAR and pharmacophore screening

A QSAR model was built using AutoQSAR (Schrödinger)[38] with the target dataset to predict the antibacterial activity. Ligand-based and e-pharmacophore hypotheses were generated using PHASE, and hits were prioritized based on fitness and PhaseScreenScore (see Supplementary Sections S1.6 and S1.7 for details).

2.3 | Molecular docking, MM-GBSA, and molecular dynamics simulations

Top-ranking compounds were docked into the DHFR active site (PDB ID: 4LAE)[39] using the Glide XP.[40,41] Following docking, 36 hits were further evaluated using a Multi-Criteria Decision Analysis (MCDA) framework. The binding energies were estimated using MM-GBSA with the Prime module.[42] Desmond (Schrödinger) was used to perform 1000 ns molecular dynamics (MD) simulations for the selected ligand–DHFR complexes (see Supplementary Sections S1.8–S1.11 for full details).[43]

2.4 | Chemical synthesis of SAK-2970

SAK-2970 was synthesized via a three-step procedure; please refer to Supplementary Section S1.12 for full details.

2.5 | Biological evaluation: Antibacterial activity, resistance profiling, in vivo efficacy and biocompatibility analysis

SAK-2970 was tested for DHFR inhibition, MIC/MBC determination, time-kill kinetics, biofilm inhibition, and eradication. Biofilm viability was visualized using CLSM and crystal violet staining techniques. Resistance development was evaluated over 30 days of sub-MIC exposure. Toxicity and efficacy were assessed in SPF Kunming and C57BL/6 mice, respectively. Bacterial loads were quantified post-infection in the blood, liver, spleen, kidneys, and lungs. Survival and organ indices were evaluated post-treatment. Hemocompatibility was assessed with human RBCs, and cytocompatibility across 293T, hMSC, and 3T3 cells using MTT assays and fluorescence microscopy (see Supplementary Sections S1.13–S1.16 for full details).

2.6 | Statistical analysis

All results are reported as mean ± standard deviation (SD). Statistical significance was determined using one-way and two-way ANOVA (GraphPad Prism), with $p < 0.05$ considered significant.

3 | RESULTS

3.1 | Establishment of the GDL model

The proposed GDL model is based on a sophisticated adaptive TargetGen-RNN architecture, wherein LSTM[44] networks are employed to model the intricate sequential dependencies intrinsic to SMILES representations. The model integrates a temperature-modulated sampling mechanism coupled with transfer learning and fine-tuning to deliver unparalleled control over scaffold diversification and target-specific optimization. These advancements confer TargetGen-RNN with superior generative capabilities, positioning it at the forefront of target-driven de novo drug discovery, particularly for the generation of biologically relevant and structurally diverse novel molecular entities. The architecture of the proposed TargetGen-RNN is illustrated schematically in Figure 1.

During pre-training, TargetGen-RNN was exposed to an extensive corpus of 5.6 million compounds (referred to as source data) (ZINC 15 database)[31] in the form of SMILES strings, facilitating the model's internalization of a broad and heterogeneous chemical space. The SMILES sequences were tokenized into individual components and subsequently mapped to integer indices using a predefined token-to-index dictionary incorporating <SOS>, <EOS>, and <PAD> tokens to ensure sequence uniformity. These token indices were then processed by an embedding layer, which transformed them into dense vector representations, thereby capturing intricate molecular relationships. The LSTM layers were meticulously optimized using categorical cross-entropy, allowing the model to capture both the local and global dependencies within the molecular structures. This exhaustive pre-training phase endowed TargetGen-RNN with profound chemical comprehension, priming it for efficient transfer learning and downstream specialization in highly specific biochemical contexts.

One of the hallmark capabilities of TargetGen-RNN lies in its proficiency in fine-tuning with remarkably small target-specific libraries, an area where traditional models, such as cRNNs, falter owing to their dependence on large data.[29] By leveraging transfer learning,[30] the model retains vast chemical knowledge from the pre-training phase and dynamically adapts it to new biological targets. For instance, fine-tuning compact target data (here, target data comprising 82 *S. aureus* DHFR inhibitors acquired from BindingDB, Supplementary Table S1)[32] allowed TargetGen-RNN to internalize critical pharmacophoric constraints, resulting in a model adept at generating inhibitors that navigate the delicate balance between novelty and biochemical efficacy, which is a challenge for other models constrained by data scarcity. It was interesting to note that TargetGen-RNN achieved high predictive accuracy (~0.90) and low convergent loss (~0.30 to 0.40), reflecting efficient learning and robust generalization (Supplementary Figure S1).

3.2 | Tailor-made generation of diverse and novel scaffold library for *S. aureus* DHFR

The aforementioned TargetGen-RNN model was employed to construct a tailored virtual compound library with a specific focus on targeting *S. aureus* DHFR. The implementation of this model resulted in the generation of 28,708 unique molecules (generated data) with redundancies, and compounds containing structural alerts or reactive functional groups were preemptively excluded. To characterize the chemical space relationship between source data, target data, and newly generated data, a UMAP plot[45] was generated. As illustrated in Figure 2A, the generated data (depicted in green) displayed a pronounced shift from the source data (blue) toward the target data (orange) following the application of transfer learning, thereby demonstrating the effectiveness of the model in traversing the chemical space from the source to the target domain.

An analysis of the relative scaffold diversity (i.e., the ratio of unique scaffolds to the total number of scaffolds)[46] revealed Murcko scaffold[47] diversities of 99.85%, 83.81%, and 89.47% for the source data, generated data, and target data, respectively (Figure 2B). Remarkably, 99.99% and 83.83% of the scaffolds within the generated data were distinct from those found in the target data and source data, respectively, resulting in an overall novelty of 83.81% (Figure 2B). Furthermore, our analysis of scaffold diversity demonstrated that the generated data not only surpassed both the target data and the source data in terms of diversity (Figure 2C,D and Supplementary Figure S2A) but also explored a broader and more unique chemical space than previously reported results.[29] Regarding fingerprint diversity,[48] the generated data were also better than the target data and closer to the source data for various types of fingerprints, including ECFP_4, MACSS_4, and PubChem _4 (Supplementary Figure S2B–G).

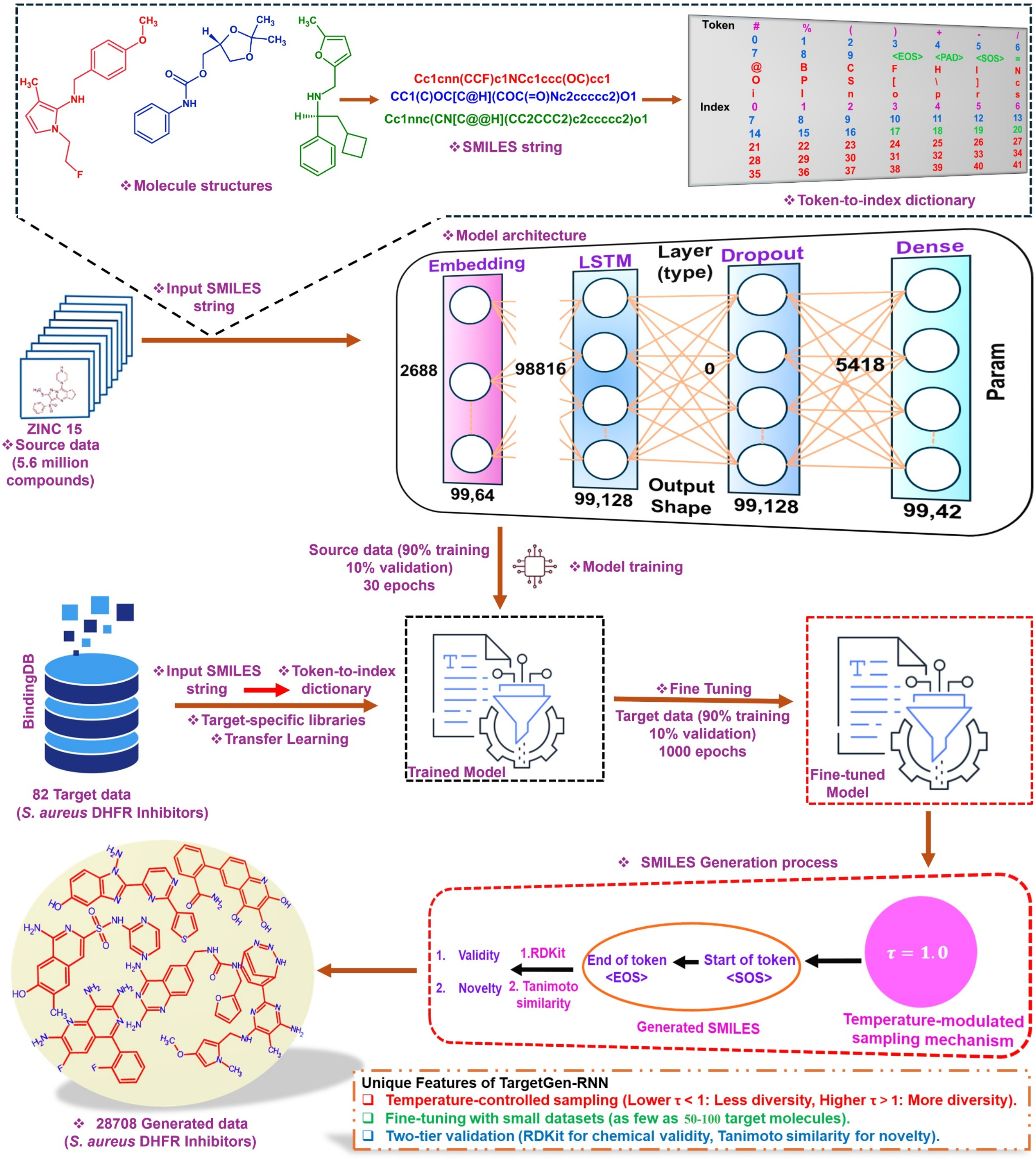

FIGURE 1 Schematic representation of the TargetGen-recurrent neural network model architecture and workflow. The model leverages a large source of data of 5.6 million compounds from the ZINC 15 database for initial training, followed by fine-tuning on a target data of 82 *Staphylococcus aureus* DHFR inhibitors. SMILES strings were tokenized and passed through the model architecture, which consisted of embedding, long short-term memory, dropout, and dense layers. The trained model was then fine-tuned with 82 smaller data, and internal validation (not shown) further suggested functional adaptation with as few as 50 compounds (see Supporting Information S1: Discussion). The temperature-controlled sampling mechanism ($\tau = 1.0$) enabled the generation of novel SMILES strings, which were validated through a two-tier process involving the RDKit for chemical validity and Tanimoto similarity for structural novelty. This model generated 28,708 unique *S. aureus* DHFR inhibitors (generated data) with verified chemical structures.

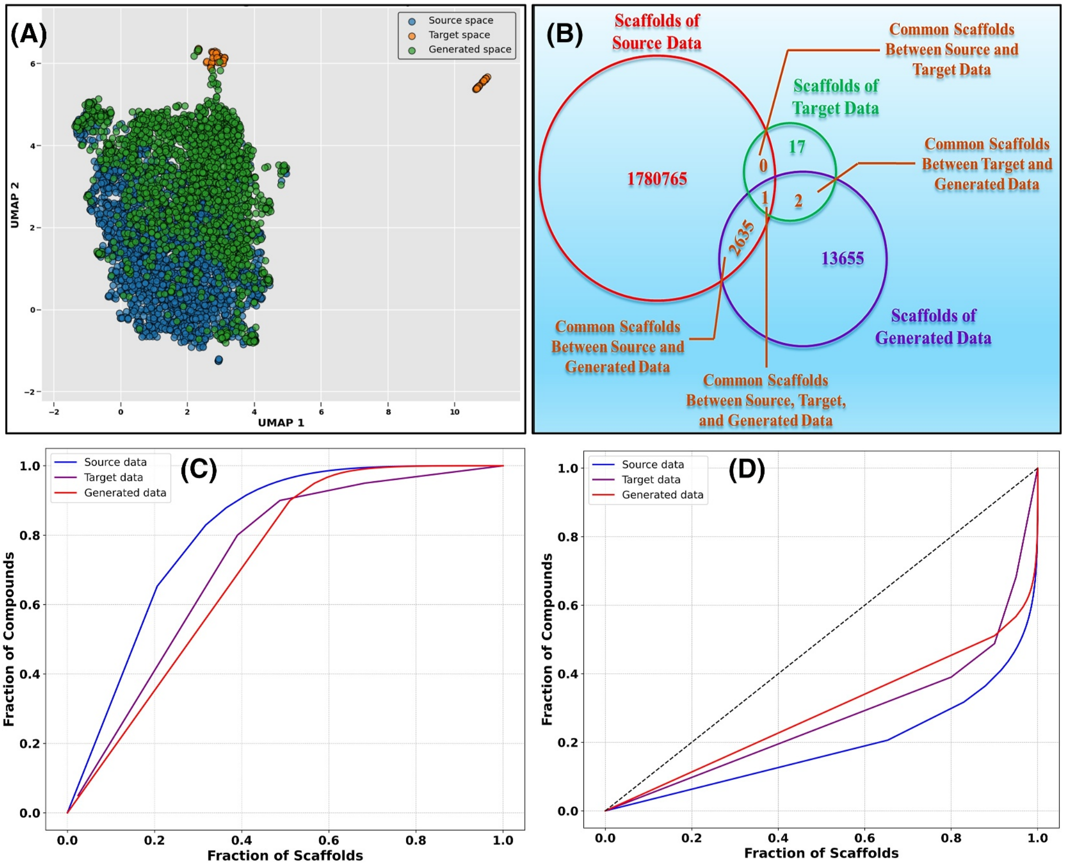

FIGURE 2 Comprehensive analysis of scaffold diversity, novelty, and chemical space transformation in the generated data compared with the source data and target data. (A) UMAP plot illustrating distinct clustering of chemical spaces: source data (blue), target data (orange), and generated data (green). The pronounced shift of the generated molecules toward the target data highlights the potential of the generative deep learning model in guiding chemical space exploration and optimizing scaffold design. (B) Venn diagram showing the generic Murcko scaffold overlap between the source data, target data, and generated data. This diagram emphasizes the novel scaffold generation capability of the TargetGen-recurrent neural network model, with a significant portion of scaffolds unique to the generated data, demonstrating its potential for expanding the chemical space for drug discovery. (C) Cumulative scaffold frequency plots and (D) Lorenz curve for the distribution of compounds over generic Murcko scaffolds in the source data, target data, and generated data. Both plots indicate the superior scaffold diversity achieved in the generated data and showcasing the model's ability to generate a broader array of scaffolds, contributing to enhanced structural diversity, which is crucial for identifying novel drug-like compounds.

3.3 | Physicochemical properties: Comparative insights and filtering outcomes

Screening of different physiochemical properties was performed to evaluate the similarity relationship between the molecules in the generated data and the target data based on them. Figure 3A and Supplementary Figure S3A–D demonstrate that the molecules in the generated data not only exhibited congruence with the active compounds of the target data but also possessed improved physicochemical properties, underscoring the capacity of the model to produce chemically and pharmacologically pertinent entities. To maintain scaffold uniqueness within the generated data, compounds with generic Murcko scaffolds or substructures identical to those of known *S. aureus* DHFR inhibitors (target data) were systematically excluded, thereby reducing the number of molecules from 28,708 to 28,410.

Figure 3B highlights the structural refinement of the generated molecules following the application of stringent drug-likeness and medicinal chemistry filters, including Lipinski's Rule, Veber's rule, Ghose's rule, QED score, PAINS, and Brenk alerts.[35–37,49] This stringent filtration reduced the generated data, narrowing the pool of molecules from 28,410 to 6672 high-quality hits. The TMAPs visualizations demonstrated that the filtered generated molecule library preserved significant chemical diversity across parameters, such as MW, LogP, TPSA, and NAR, affirming the model's ability to generate compounds that conform to classical drug-likeness criteria while maintaining structural novelty. The

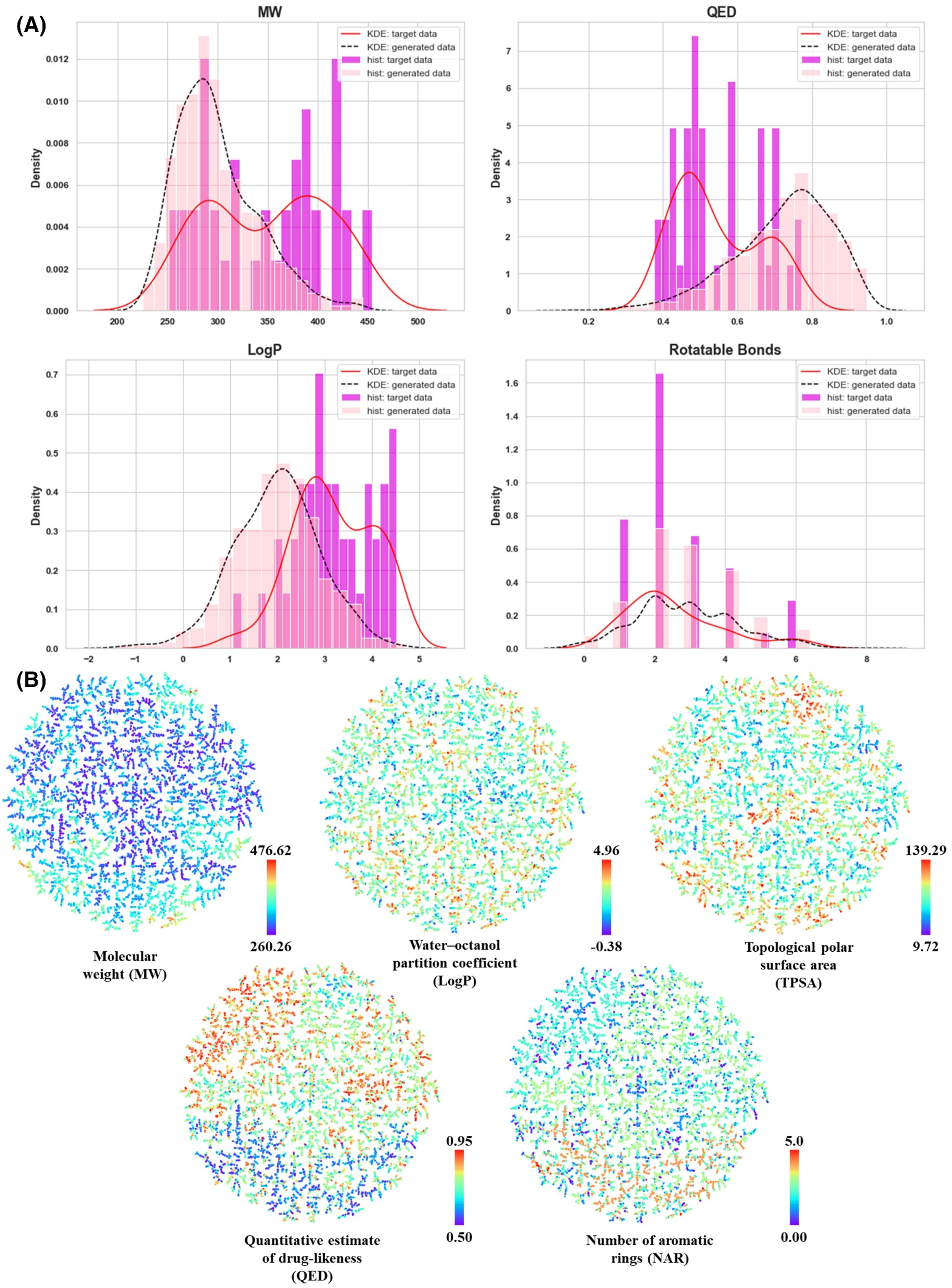

FIGURE 3 Comparative analysis of physicochemical properties between the target and generated data and post-filtering visualization. (A) Histograms and Kernel Density Estimation (KDE) plots comparing key physicochemical properties, including molecular weight (MW), quantitative estimate of drug-likeness (QED), water-octanol partition coefficient (LogP), and number of rotatable bonds across the target data and generated data. The plots revealed enhanced diversity and favorable distribution of properties in the generated data, closely aligning with the desired physiochemical profile of the target data. (B) Visualization of five key physicochemical properties of the generated molecules after applying filters based on Lipinski's Rule, Veber's rule, Ghose's rule, QED score, SA score, PAINS, and Brenk alerts. TMAPs highlight the structural diversity and compliance with drug-likeness rules within the filtered compound library.

retention of diversity post-filtering is particularly critical, as it indicates that the generated molecules not only adhere to essential medicinal chemistry principles but also maintain a variety of scaffolds necessary to mitigate the risk of attrition during drug development. The precisely modulated TPSA values suggest an optimized hydrogen bonding potential, thereby enhancing target specificity while reducing off-target effects. Moreover, the range of NAR values indicates the generation of aromatic systems that are conducive to π–π stacking interactions, which are crucial for stabilizing the drug-target complexes.

This analysis underscores the advanced capability of the model to generate a compound library that achieves an optimal balance between chemical innovation and adherence to medicinal chemistry. The generated data stands as a promising resource for the identification of novel scaffolds with the potential for high efficacy, favorable pharmacokinetics, and reduced toxicity, thereby providing a solid foundation for subsequent lead optimization and preclinical development.

3.4 | Antibiotic activity profiling of generated compounds

To evaluate the antibiotic potential of the generated compounds, a series of QSAR models were developed using the AutoQSAR module in the Schrödinger Maestro suite.[38] Ten models were initially constructed, and the key performance metrics are summarized in Supplementary Table S2. Among the 10 models constructed, mlr_33 demonstrated superior performance, as evidenced by an R^2 value of 0.9378 and a low RMSE of 0.2303, underscoring its predictive accuracy and robustness. The model's Q^2 value of 0.9398 further confirmed its reliability through rigorous cross-validation.

The performance of the model mlr_33 was comprehensively analyzed using several key plots (Figure 4A–C). The observed versus predicted activity scatter plot revealed a strong correlation between the observed and predicted values across both the training and test data, with data points clustering tightly along the $y = x$ line. This close alignment indicates the precision of the model in predicting the antibiotic activity with a minimal risk of overfitting or underfitting. The residual plot displayed a symmetric distribution of residuals around zero, suggesting that the prediction errors were random and unbiased, further validating the model reliability across a diverse set of compounds. The QQ plot of the residuals indicated that they followed an approximately normal distribution with only minor deviations at the extremities. These slight deviations, indicative of a few outliers, did not significantly compromise the model's overall accuracy and predictive capability.

Subsequent QSAR analysis of the 6639 generated compounds using Model mlr_33 provided a detailed assessment of their predicted antibiotic activity, which was visualized through a UMAP plot and a prediction score distribution graph. In the UMAP plot (Figure 4D), the compounds were clustered based on their predicted pIC_{50} values and categorized into five distinct groups. Compounds with high predicted activity (pIC_{50} values of 12–15) are represented by red dots, whereas those with moderate activity levels (pIC_{50} values between 9–11.99 and 6–8.99) are denoted by blue and green dots, respectively. Compounds with lower predicted activity (pIC_{50} values between 3 and 5.99 and below 3) are indicated by yellow and brown dots. The cumulative distribution of the predicted pIC_{50} scores (Figure 4E) revealed that 52.3% of the generated compounds, equating to 3472 hit compounds, exceeded the average pIC_{50} threshold of 7.37, as highlighted by the red line. Compounds that surpass this threshold are expected to exhibit significant antibiotic activity, with a subset achieving pIC_{50} scores of 12 or higher, suggesting a strong potential for potent antibiotic effects. These results highlight the antibiotic potential of the generated compounds, with 52.3% surpassing the average pIC_{50} threshold, positioning them as strong candidates for further experimental validation against MDR *S. aureus*.

3.5 | Pharmacophore-based virtual screening

Building upon the AutoQSAR analysis, we employed two complementary pharmacophore-based virtual screening strategies: ligand-based and receptor-based approaches. In ligand-based pharmacophore screening, we constructed an exhaustive pharmacophore model derived from the target data of 82 known DHFR inhibitors, which yielded 14 unique pharmacophore hypotheses (Supplementary Table S3). Of these, four hypotheses (designated DDRRR_1 to DDRRR_4) comprised five features: two hydrogen bond donors (D) and three aromatic ring systems (R), whereas the remaining 10 hypotheses (DDRR_1 to DDRR_10) contained four features: two hydrogen bond donors and two aromatic ring systems. Hypothesis DDRRR_1 (Figure 4F) emerged as the top performer, as evidenced by its superior Phase Hypo Score, underscoring its efficacy in identifying potent hits during virtual screening.[50] Furthermore, the enrichment plot of DDRRR_1 (Figure 4G) substantiates its robustness, demonstrating its ability to rapidly retrieve nearly 100% of the active compounds within the top-ranked subset of

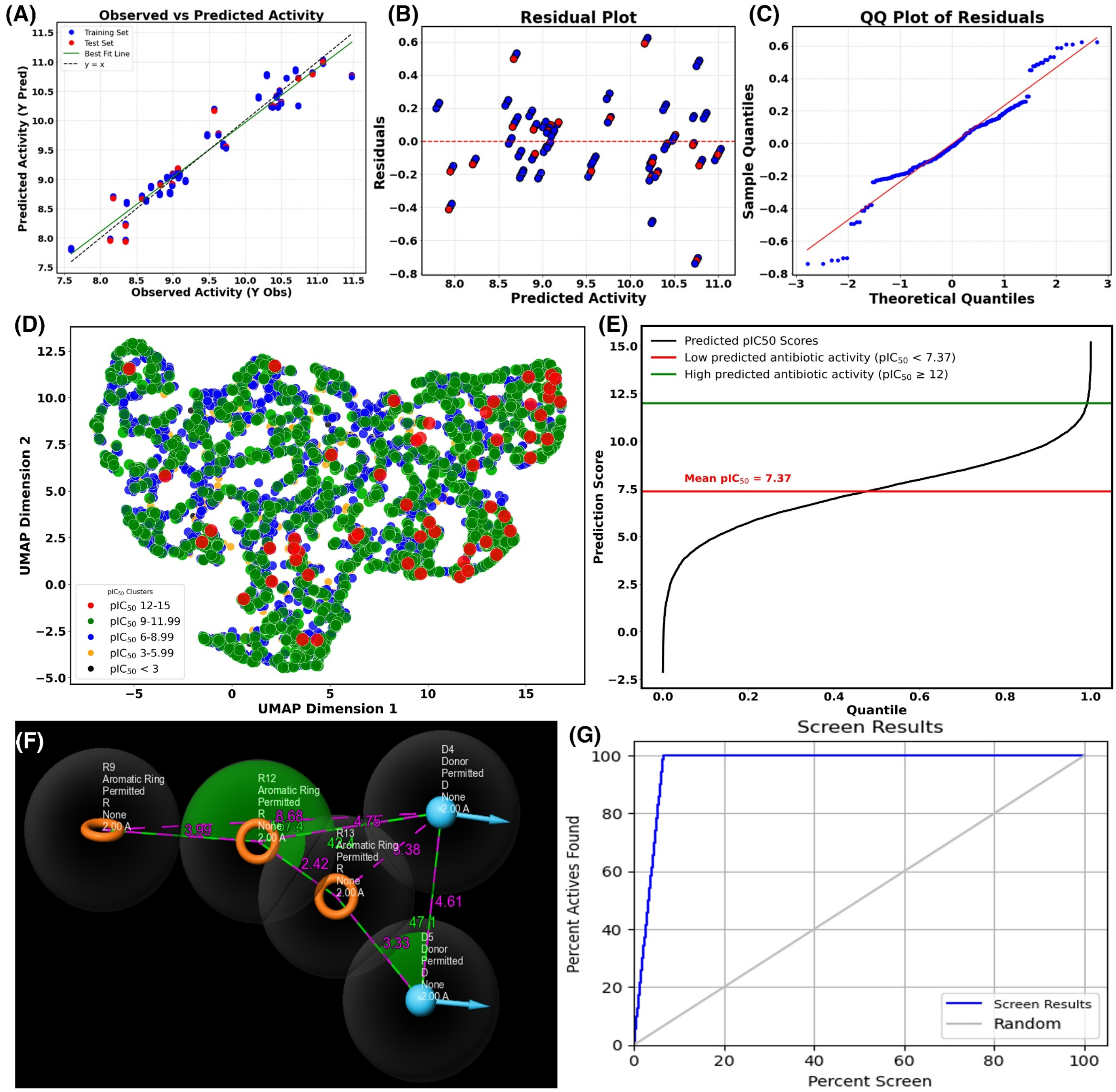

FIGURE 4 Comprehensive analysis of the predictive performance of the QSAR model and antibiotic activity profiling of the generated compounds. (A) Scatter plot of the predicted versus observed activity for the training and test sets using the mlr_33 model. (B) Residual plot showing residuals symmetrically distributed around zero, confirming the model's reliability. (C) QQ plot indicating that the residuals generally followed a normal distribution with minimal outliers. (D) UMAP clustering of compounds by predicted pIC_{50} values, highlighting groups with high (pIC_{50} 12–15), moderate (pIC_{50} 9–11.99, 6–8.99), and low (pIC_{50} 3–5.99, <3) predicted antibiotic activity. (E) Cumulative distribution of predicted pIC_{50} scores, with 52.3% of compounds exceeding the pIC_{50} threshold of 7.37, suggesting outstanding antibiotic potential. (F) Ligand-based pharmacophore hypothesis: the figure illustrates the top pharmacophore hypothesis, DDRRR_1, comprising two hydrogen bond donors (D4 and D5) and three aromatic ring systems (R9, R12, and R13). (G) Enrichment plot of the pharmacophore hypothesis DDRRR_1. The plot shows the percentage of actives found as a function of the percent screen for the top pharmacophore hypothesis, DDRRR_1.

screened data. This indicates that DDRRR_1 is a highly reliable and efficient model for virtual screening, thus serving as a valuable tool for prioritizing potential drug candidates. Subsequently, the generated data of 3472 hit compounds were screened and filtered for compounds matching at least four features of the pharmacophore model with a fitness and PhaseScreenScore ≥1.5, thereby narrowing the hit pool to 553 compounds. Additionally, the superimposition of the top 20 hits onto the DDRRR_1 hypothesis (Supplementary Figure S4) highlights the critical interactions delineated by the pharmacophore features, further validating their potential as potent inhibitors.

To further refine the selection and identify robust and highly active hits among the 553 compounds, receptor-based pharmacophore screening was conducted. A comprehensive pharmacophore model was constructed based on the co-crystal structure of the DHFR–ligand complex (Figure 5A), which encapsulated all critical interactions between DHFR and its ligands. This pharmacophore model (designated DDDRRR) included six features: three hydrogen bond donors (D3, D4, and D6) and three aromatic rings (R13–R15) (Figure 5B). Compounds that satisfied the stringent criterion of matching at least five pharmacophore features, coupled with a fitness and PhaseScreenScore threshold of ≥1.5, were meticulously selected, as illustrated in Figure 5C. This rigorous screening process culminated in the identification of 74 highly potent compounds. The superimposition of the top 20 hits onto the DDDRRR hypothesis is presented in Supplementary Figure S5. Notably, these compounds exhibited robust interaction profiles with DHFR, while preserving the scaffold and fingerprint diversity (Figure 5D) inherent in the original generated data. The clustering of these high-fitness compounds within a narrow range of PhaseScreenScore underscores the precision of the screening process, reinforcing the potential of these hits as exceptional druggable candidates for subsequent experimental validation.

3.6 | Molecular docking and MCDA analysis

Molecular docking was performed to prioritize the 74 hit compounds identified in a previous pharmacophore-based virtual screening. Figure 6A presents the docking scores (kcal/mol) epitomized on the t-SNE plot via a color-coded gradient, where blue hues correspond to more favorable binding interactions (lower docking scores), whereas red hues signify less favorable interactions (higher docking scores). Notably, 36 compounds exhibited docking scores > −7.270 kcal/mol (average docking score). The chemical structures of these compounds and their docking scores are presented in Supplementary Figure S6. Moreover, the results showed that the reference ligand (1VM) demonstrated a docking score of −8.098 kcal/mol, and nine compounds surpassed this benchmark, indicating a stronger binding affinity for the DHFR active site. The three compounds (SAK-2970, SAK-2257, and SAK-5557) with the highest docking scores are shown as docked complexes in Figure 6B–D.

The implementation of MCDA subsequently refined and prioritized the 36 compounds identified as potential therapeutic agents. MCDA ranked the seven best compounds (SAK-2970, SAK-3702, SAK-4367, SAK-6129, SAK-4409, SAK-2038, and SAK-1893) (Figure 6E) based on an integrated assessment of pharmacokinetic parameters, physicochemical properties, pIC_{50} values, docking scores, and ADMET profiles. These compounds demonstrated MCDA scores ranging from 0.46–0.58. The most promising compounds exhibited an optimal balance between high binding affinity and potent biological activity, suggesting significant therapeutic potential. Notably, none of the most promising compounds demonstrated a significant level of inhibition of the hERG channel, thus mitigating the potential for cardiotoxicity. Furthermore, these compounds demonstrated favorable bioavailability (Fub_human) and high SA, facilitating their development and scalable synthesis. Their compliance with the QED further corroborates their potential for medicinal chemistry optimization. However, each of the remaining 29 compounds exhibited unfavorable properties in one or more parameters, thereby reducing their suitability as potential therapeutic candidates.

3.7 | MM-GBSA and MD simulation analysis

To further evaluate the stability of DHFR-docked complexes with SAK-2970, SAK-4367, SAK-4409, SAK-3702, SAK-6129, SAK-1893, and SAK-2038, their ΔG_{Bind} values were calculated. Notably, all complexes demonstrated negative ΔG_{Bind} values, indicating the formation of robust and stable DHFR-compound binding (Supplementary Table S4). These findings corroborate that the compounds establish energetically favorable interactions with the DHFR active site, highlighting their potential for stable and sustained binding.

MD simulations over 1000 ns were subsequently performed to evaluate the stability and binding dynamics of the DHFR complexes with the six generated compounds relative to the reference ligand (Figure 7A–E; Supplementary Figures S7–S9A–F; Supplementary Table S5). The protein backbone RMSD indicated that all complexes were equilibrated within 800 ns. The SAK-2970 complex exhibited fluctuations nearly identical to those of the reference ligand (≈1.4–1.6 Å), whereas SAK-4409 and SAK-2038 remained close to the reference range. In contrast, SAK-3702 showed the largest deviation (2.15 ± 0.28 Å), suggesting weaker protein stabilization.

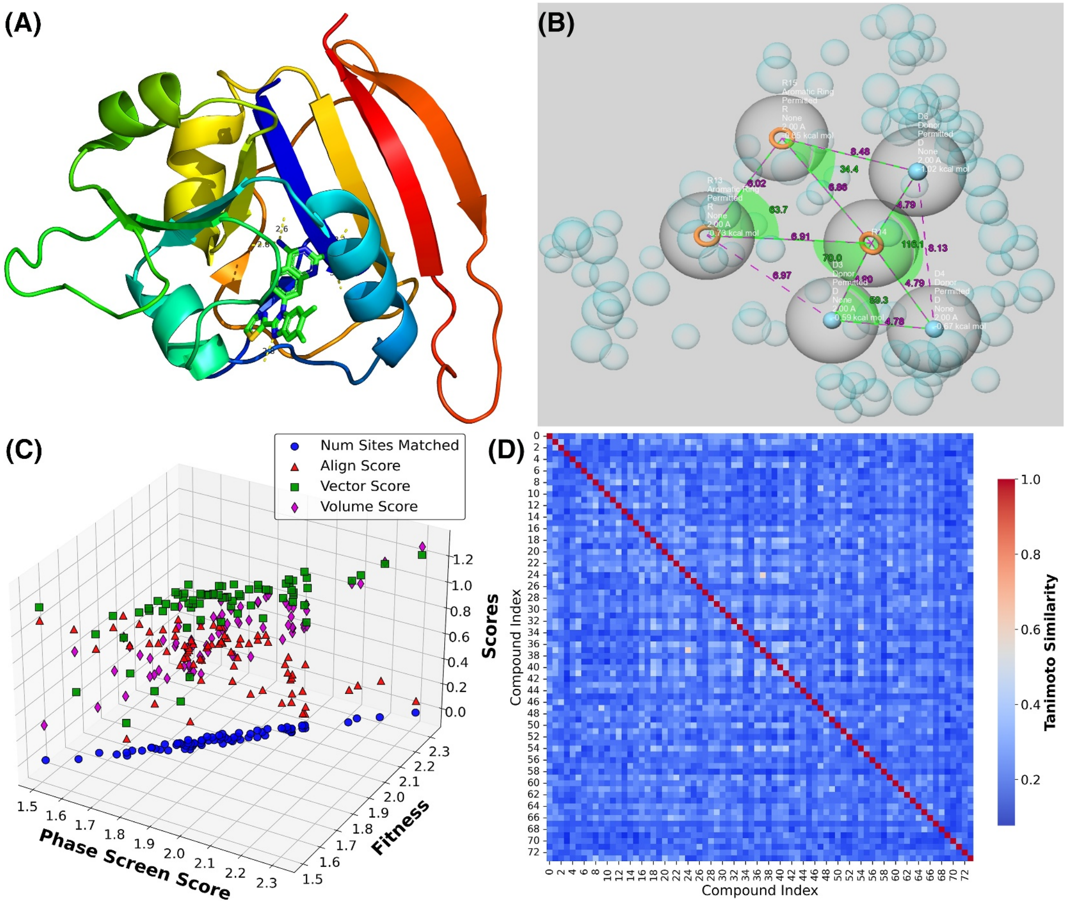

FIGURE 5 (A) Co-crystal structure of the DHFR-ligand complex. (B) The receptor-based pharmacophore hypothesis comprised six features: three hydrogen bond donors (D3, D4, and D6) and three aromatic rings (R13–R15). (C) 3D scatter plot of fitness versus phase screen score: This 3D scatter plot offers a comparative analysis of the fitness and PhaseScreenScore of the 74 hit compounds against various pharmacophore metrics, including the number of matched sites, alignment score, vector score, and volume score. The distribution and clustering of points reveal compounds with optimal alignment and fitness, which are critical for prioritizing the most promising candidates. (D) Heatmap of pairwise Tanimoto similarities: The heatmap visualizes the pairwise Tanimoto similarities among the 74 hit compounds using a color gradient from blue to red to indicate increasing similarity. The diagonal represents perfect self-similarity, whereas the overall color distribution highlights the structural diversity preserved within the hit compounds, as inherent in the original generated data.

The protein-RMSF analysis showed that the reference ligand complex had the lowest overall flexibility, with SAK-2970 closely matching this profile. SAK-3702 induced elevated fluctuations in several regions, whereas the other candidates exhibited intermediate behaviors.

Interaction analysis (Figure 7C; Supplementary Figures S7 and S9E,F) revealed that the SAK-2970, SAK-2038, SAK-1893, and reference ligands formed multiple direct and water-mediated hydrogen bonds as well as π–π stacking interactions, contributing to strong stabilization. In contrast, SAK-3702, SAK-4367, and SAK-4409 engaged in fewer stabilizing Ligand RMSD values further emphasized these differences (Figure 7D; Supplementary Figure S9C, Supplementary Table S5): the reference ligand (0.73 ± 0.06 Å) was closely matched by SAK-2970 (0.72 ± 0.04 Å), while SAK-2038 showed moderate deviations (1.51 ± 0.54 Å). By contrast, SAK-1893 exceeded 5 Å, reflecting weak or transient binding. SAK-3702, SAK-4367, and SAK-4409 also displayed high RMSD values (4.31–4.61 Å), suggesting substantial ligand mobility and reduced binding stability.

Ligand RMSF analysis (Supplementary Figure S8) confirmed minimal fluctuations for SAK-2970 and SAK-4367, modest variation for SAK-2038, and pronounced flexibility for SAK-1893, SAK-3702, and SAK-4409. Furthermore, rGyr values were stable across all systems (Supplementary Table S5), with the reference ligand at 4.13 Å and all generated complexes slightly more compact (≈3.40–3.98 Å), although this compaction did not offset the higher ligand mobility observed in the weaker binders. Collectively, these analyses demonstrated that SAK-2970 most closely reproduces the reference ligand-binding

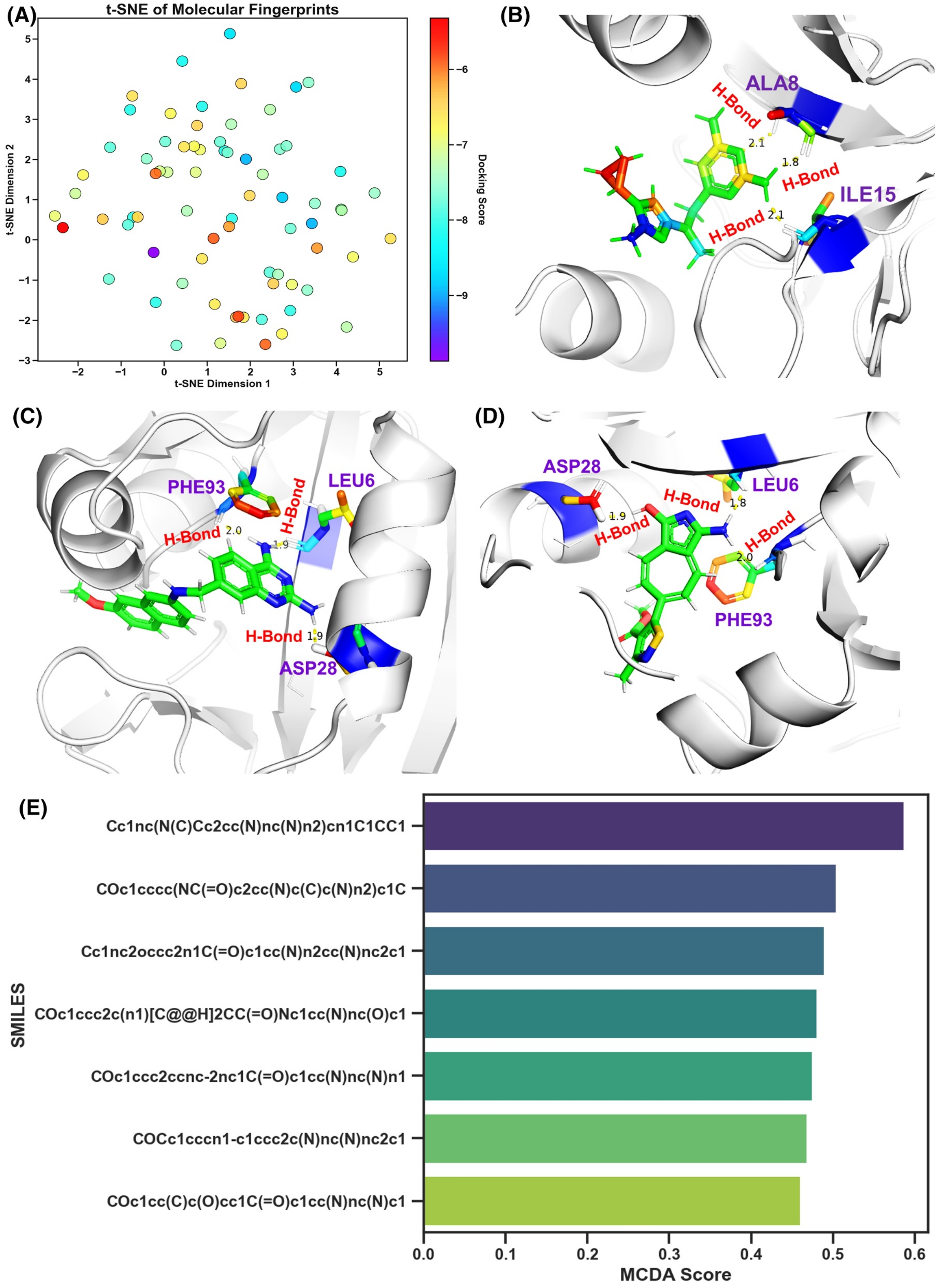

FIGURE 6 (A) t-SNE visualization of molecular fingerprints. The t-SNE plot shows the molecular diversity of the generated data. Color-coded by docking scores, it highlights regions of interest where higher binding affinities (in warmer colors) cluster, indicating more promising docking results. (B) Compound (SAK-2970) binds to the critical residues ALA8 and ILE15 of DHFR through hydrogen bonds. (C) Compound (SAK-2257) bound to residues ASP28, LEU6, and PHE93 through multiple H-bonds. (D) Compound (SAK-5557) bound to the active sites through hydrogen bonds with ASP28, LEU6, and PHE9. (E) The compounds (SAK-2970, SAK-3702, SAK-4367, SAK-6129, SAK-4409, SAK-2038, and SAK-1893) were ranked according to their MCDA scores (0.586, 0.504, 0.489, 0.480, 0.474, 0.468, and 0.460), with higher scores indicating more promising therapeutic drug candidates.

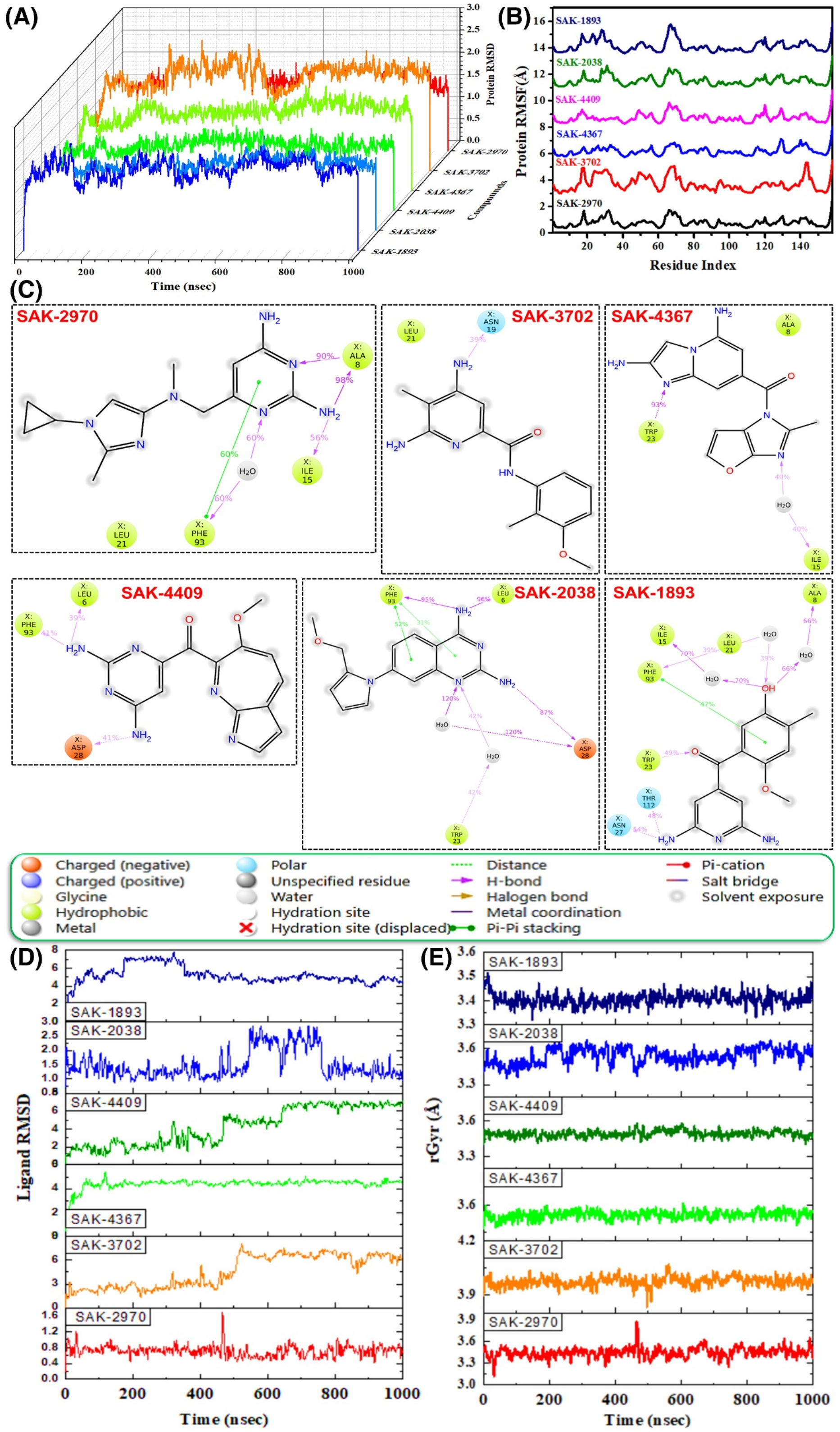

FIGURE 7 Molecular dynamics simulation results for the generated compound–DHFR complexes. (A) RMSD of the Cα backbone of DHFR bound to the generated compounds. (B) Cα backbone RMSF of DHFR bound to the generated compounds. (C) Ligand-protein contacts. (D) RMSD of generated compounds bound to DHFR. (E) Radius of gyration of the generated compounds bound to DHFR.

profile across all MD metrics, SAK-2038 exhibits intermediate stability, and the remaining compounds are destabilized by high ligand fluctuations despite reasonable protein-level stability.

3.8 | Synthesis of SAK-2970: A potent antibiotic compound

SAK-2970 was chemically synthesized according to the scheme presented in Figure 8A. The general procedure is as follows (for details, please see Supplementary Information S1): commercially available (**1**) was reacted with thionyl chloride (**2**) to give intermediate (**3**), which was aminated and reacted with methylamine to afford intermediate (**4**). Intermediate (**4**) was reacted with 4-bromo-1-cyclopropyl-2-methyl-1H-imidazole (**5**) via the Buchwald–Hartwig amination reaction to produce compound SAK-2970.

3.9 | Antibacterial performance of SAK-2970

The synthesized compound SAK-2970 was evaluated for its biological performance. Initially, its inhibitory activity against DHFR was examined using a previously reported protocol.[51] The results demonstrated that SAK-2970 exhibited potent inhibitory activity against DHFR, with an IC_{50} of 2.379 nM (Figure 8B). In contrast, methotrexate showed significantly lower inhibition, with an IC_{50} of 0.12 μM, which is in close agreement with reported results.[52] Subsequently, the antibacterial efficacy of SAK-2970 was evaluated by determining its MIC and MBC against *S. aureus*. The results demonstrated that SAK-2970 not only effectively inhibited the growth of *S. aureus* but also exhibited significant bacteriostatic activity, with an MIC/MBC value of 13.48 ± 2.71 μg/mL (Figure 8C). Bacterial growth inhibition was also assessed by measuring the optical density at 600 nm, revealing that SAK-2970 exhibited potent and markedly superior bactericidal activity compared to the standard drug ciprofloxacin (Figure 8D).

To elucidate the temporal dynamics of bacterial eradication, we evaluated the time-dependent bactericidal activity of SAK-2970. At the MIC/MBC dose, SAK-2970 demonstrated a rapid, time-dependent bactericidal effect against *S. aureus*, eradicating over 60% of the bacterial population within 8 h, thus surpassing the efficacy of ciprofloxacin (Figure 8E). The inhibition of bacterial growth increased proportionally over time, with no visible growth of *S. aureus* observed in the media after 24 h of treatment with SAK-2970 (Figure 8F). In contrast, ciprofloxacin exhibited markedly lower inhibitory activity (Supplementary Figure S10A). This effect was further corroborated by plating the inoculum on LB agar plates, where an approximately 92% reduction in the CFUs of *S. aureus* was observed after 8 h of incubation with SAK-2970, culminating in >99.998% inhibition after 24 h of treatment. Conversely, ciprofloxacin exhibited a significantly lower bactericidal potential than SAK-2970 (Figure 8G). Similarly, live/dead staining further substantiated that after a 24-h incubation with SAK-2970, *S. aureus* exhibited a markedly higher proportion of dead cells, as indicated by red fluorescence, when compared to ciprofloxacin treatment (Figure 8H).

3.10 | Antibiofilm efficacy

Bacterial biofilms constitute intricate ecosystems, composed of diverse bacterial colonies that aggregate into dense clusters and exhibit significant resistance to antibiotics and chemical agents. This resilience leads to persistent bacterial infections, which present substantial challenges in healthcare, medical, and industrial contexts. Individuals with underlying chronic conditions that compromise the epidermal barrier are particularly susceptible to tissue-associated biofilm infections.[53] Therefore, the antibiofilm efficacy of SAK-2970 was meticulously evaluated against *S. aureus* following a 24 h incubation period, in comparison to ciprofloxacin. The results showed that SAK-2970, at ½ MIC and 1 MIC, exhibited exceptional antibiofilm activity by inhibiting biofilm formation by >65% and >90%, respectively, relative to ciprofloxacin administered at equivalent dosages (Figure 9A,B). This pronounced inhibitory effect was further substantiated by crystal violet staining (Figure 9C). To further substantiate the antibiofilm efficacy of SAK-2970, confocal laser scanning microscopy (CLSM) was used to examine *S. aureus* biofilm formation at the MIC of SAK-2970 and ciprofloxacin. These results unequivocally demonstrate that SAK-2970 profoundly inhibited the formation of *S. aureus* biofilms. Although ciprofloxacin also exhibited an inhibitory effect, its efficacy was markedly inferior to that of SAK-2970 (Figure 9D). These findings are in concordance with the assessments of optical density, percentage of biofilm inhibition, and crystal violet staining, thereby reinforcing the superior antibiofilm performance of SAK-2970.

To further validate the antibiofilm performance, biofilm eradication was quantified by measuring the optical density following incubation of preformed *S. aureus* biofilms with SAK-2970 and ciprofloxacin at concentrations of 2 × MIC. As shown in Figure 9E,F, SAK-2970 exhibited outstanding antibiofilm activity, achieving >85% biofilm eradication. In contrast, ciprofloxacin also

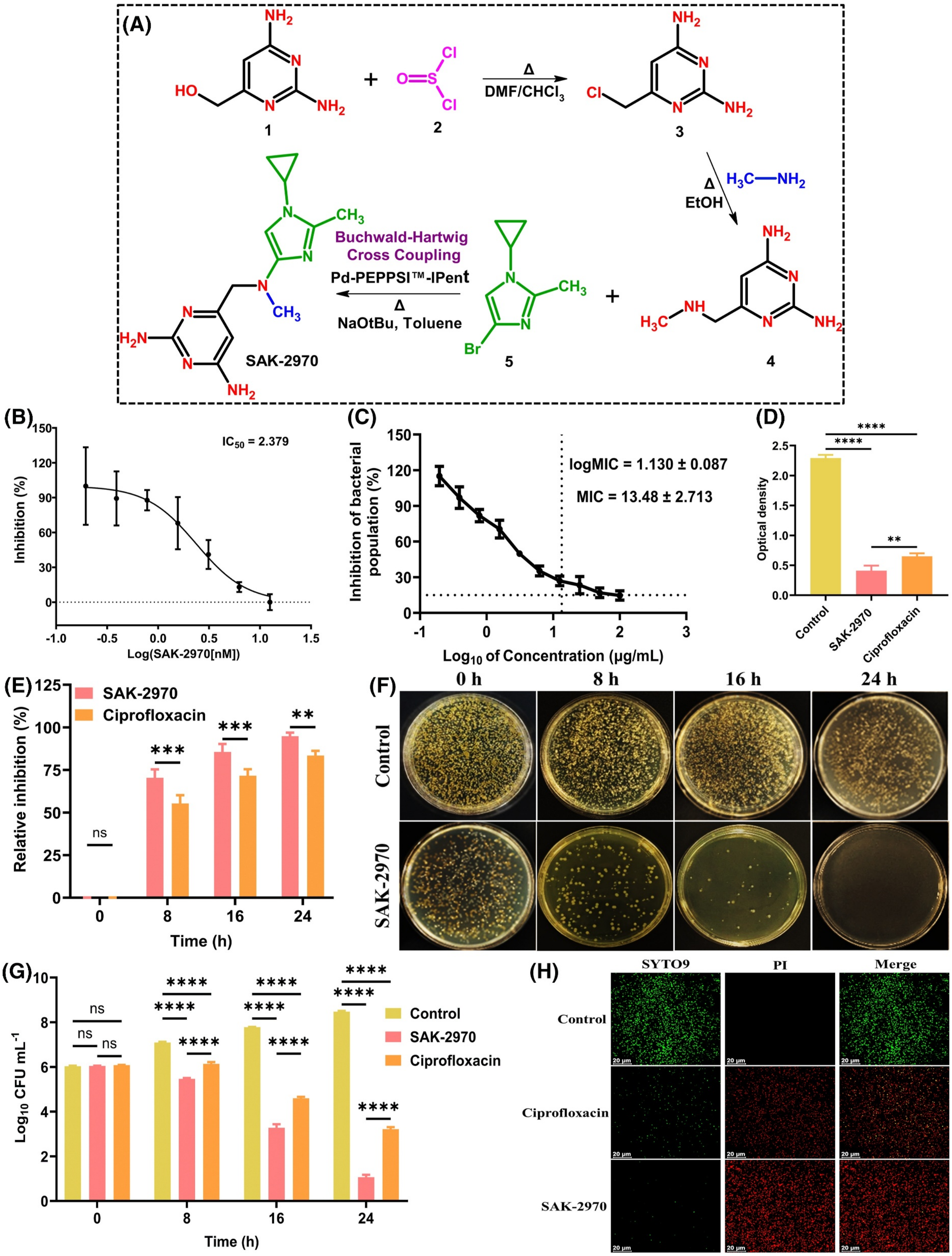

FIGURE 8 (A) Schematic illustration of the synthetic pathway for SAK-2970. (B) Dose-response curve depicting the inhibitory activity of SAK-2970 against DHFR. (C) Dose-dependent antibacterial efficacy of SAK-2970 against *Staphylococcus aureus*, demonstrating MIC and MBC. (D) Comparative analysis of bacterial growth inhibition of *S. aureus* by SAK-2970 and ciprofloxacin versus the control, as represented by optical density measurements (**$p < 0.01$; ****$p < 0.0001$). (E) Temporal bactericidal dynamics of SAK-2970 and ciprofloxacin, illustrating time-dependent bacterial reduction (ns = $p > 0.5$, **$p < 0.005$, and ***$p < 0.0005$). (F) Representative colony images of *S. aureus* treated with SAK-2970 and the control over time. (G) Quantitative assessment of the antibacterial performance of SAK-2970 relative to the control and ciprofloxacin, expressed as a reduction in Log_{10} CFU/mL across different time intervals. (H) Live/dead fluorescent staining of *S. aureus* using SYTO9/PI, highlighting viability following treatment with SAK-2970 and ciprofloxacin (ns = $p > 0.05$; ****$p < 0.0001$) (scale bar = 20 μm). Data are expressed as mean ± SD of three independent experiments.

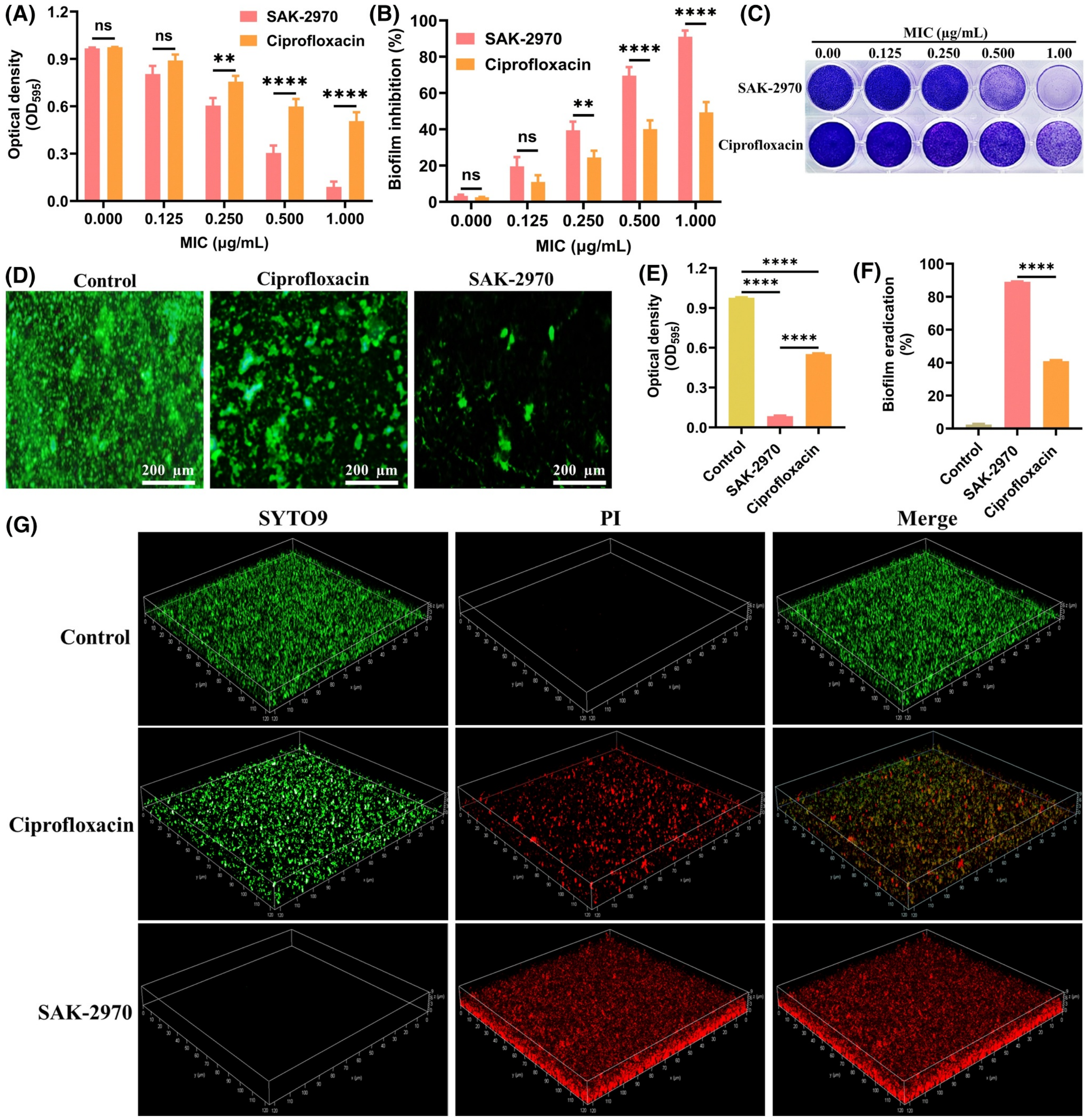

FIGURE 9 Antibiofilm efficacy of SAK-2970 against *Staphylococcus aureus* compared with ciprofloxacin. Inhibition of *S. aureus* biofilm formation by SAK-2970 at various sub-inhibitory MICs compared to ciprofloxacin, assessed by (A) optical density measurements, (B) biofilm inhibition percentage, (C) crystal violet staining assay, and (D) confocal laser scanning microscopy (CLSM). Evaluation of biofilm eradication by SAK-2970 at 2 × MIC against *S. aureus* relative to ciprofloxacin at the same concentration, as demonstrated by (E) OD measurements, (F) biofilm eradication percentage, and (G) three-dimensional CLSM live/dead imaging (scale bar = 5 µm). Data are expressed as the mean ± SD of three independent experiments (ns = $p > 0.05$, $**p < 0.005$, and $****p < 0.0001$).

demonstrated biofilm-eradicating effects (>41%); however, its efficacy was markedly inferior to that of SAK-2970. Given that the destabilization of cell membrane integrity exerts detrimental effects on the bacterial population within biofilms,[53] the impact of SAK-2970 on the cell membranes of biofilm-associated *S. aureus* was further investigated using CLSM in conjunction with SYTO9/PI dyes. As depicted in Figure 9G, CLSM images revealed that the control biofilms comprised thick, uniformly distributed green fluorescent clusters, indicating viable *S. aureus* with intact cell membranes. In stark contrast, SAK-2970-treated biofilms exhibited extensive red fluorescence, indicating complete bacterial cell death across all biofilm layers. Conversely, ciprofloxacin-treated biofilms displayed a mixture of green and red fluorescence, reflecting the presence of both viable and non-viable cells, thereby demonstrating comparatively modest antibiofilm efficacy, as previously reported.[54] These findings substantiate that SAK-2970 exhibits exceptional antibiofilm efficacy against *S. aureus*, markedly outperforming ciprofloxacin in both biofilm inhibition and eradication. This underscores its potential as a superior therapeutic agent for persistent biofilm-associated infections.

3.11 | Safety and biocompatibility assessment

For the meticulous evaluation of novel pharmacological agents, establishing hemocompatibility and cytocompatibility is imperative to ascertain their clinical viability. The safety and biocompatibility assessment of SAK-2970 compared with ciprofloxacin revealed promising results for SAK-2970's tolerability profile. In the acute toxicity evaluation (Figure 10A), SAK-2970 maintained a stable body weight in mice over 14 days at a high dose (1500 mg/kg), with significantly less weight loss than ciprofloxacin ($^{****}p < 0.0001$). Organ-specific indices further underscored SAK-2970's biocompatibility; the spleen index (Figure 10B) showed no significant difference ($p > 0.281$) between the PBS- and SAK-2970-treated groups, in contrast to the notable reduction observed in the ciprofloxacin group ($^{**}p = 0.001$). Similarly, in the thymus index (Figure 10C), SAK-2970 showed no adverse effects, whereas ciprofloxacin treatment led to marked thymic atrophy ($^{****}p < 0.0001$). Hematological profiles indicated a stable WBC count in SAK-2970-treated mice (Figure 10D), with significant reductions observed only in the ciprofloxacin group ($^{****}p < 0.0001$), and RBC counts (Figure 10E) remained unaffected by SAK-2970 compared to ciprofloxacin ($^{***}p = 0.0002$). Collectively, these findings suggest that SAK-2970 demonstrates superior safety and biocompatibility with minimal systemic toxicity and reduced hematological perturbations compared to ciprofloxacin.

To further assess the biocompatibility of SAK-2970, a hemolysis assay was performed. As illustrated in Figure 10F,G, exposure of RBCs to SAK-2970 yielded a hemolysis percentage of 3.90 ± 0.10%, which was markedly lower than that induced by ciprofloxacin. This minimal hemolytic effect demonstrates SAK-2970's benign interaction with RBC membranes, indicating its low cytotoxic profile. Moreover, a similar hemolysis percentage for ciprofloxacin has also been reported previously.[55] In stark contrast, RBCs exposed to Triton X-100, the positive control, exhibited complete rupture, resulting in an intensely red solution characteristic of significant hemolysis. These findings align with the international hemocompatibility standards for pharmaceutical excipients, which stipulate that hemolysis values <10% are indicative of non-hemolytic substances, whereas values >25% indicate high hemolytic potential.[56] Thus, SAK-2970 fulfills the stringent criteria for hemocompatibility, underscoring its promise as a safe therapeutic candidate.

The cytocompatibility of SAK-2970 was rigorously assessed in three normal cell lines: 293T human embryonic kidney cells, human mesenchymal stem cells (hMSCs), and NIH/3T3 mouse fibroblasts. As shown in Figure 10H, SAK-2970 maintained a cell viability above 83% at its MIC across all tested cell lines, a result comparable to that of ciprofloxacin, with no statistically significant differences ($p > 0.05$). It is well known that a sample with cell survival viability >75% was considered to be non-cytotoxic.[57] These findings substantiate the fact that SAK-2970 does not impart appreciable toxicity to cellular proliferation. To further validate these results, comprehensive analyses were conducted using CLSM for 293T cells and optical microscopy for hMSCs and 3T3 cells, as depicted in Figure 10I–K, respectively. CLSM analysis showed that SAK-2970- and ciprofloxacin-treated cells maintained a morphology similar to that of the control cells, with minimal PI staining, indicating intact membrane integrity. Similarly, optical microscopy images showed that both hMSCs and 3T3 cells treated with SAK-2970 retained their characteristic spindle-shaped morphology and demonstrated normal cellular architecture, comparable to that of the control cells and those treated with ciprofloxacin. These observations underscore the excellent cytocompatibility of SAK-2970, with no discernible cytotoxic effects relative to ciprofloxacin, highlighting its suitability as a biocompatible therapeutic agent.

3.12 | In vivo antibacterial performance

To further evaluate the in vivo therapeutic efficacy of SAK-2970, a murine sepsis model was established by intraperitoneal injection of *S. aureus* to investigate the pharmacological effects of the compound on bacterial infection. The results demonstrated that SAK-2970 exhibited remarkable bactericidal activity and provided a significant survival advantage over ciprofloxacin. Quantitative analysis of bacterial burden across major organs demonstrated significant reductions in CFUs within the blood, liver, spleen, kidneys (Figure 11A–D),

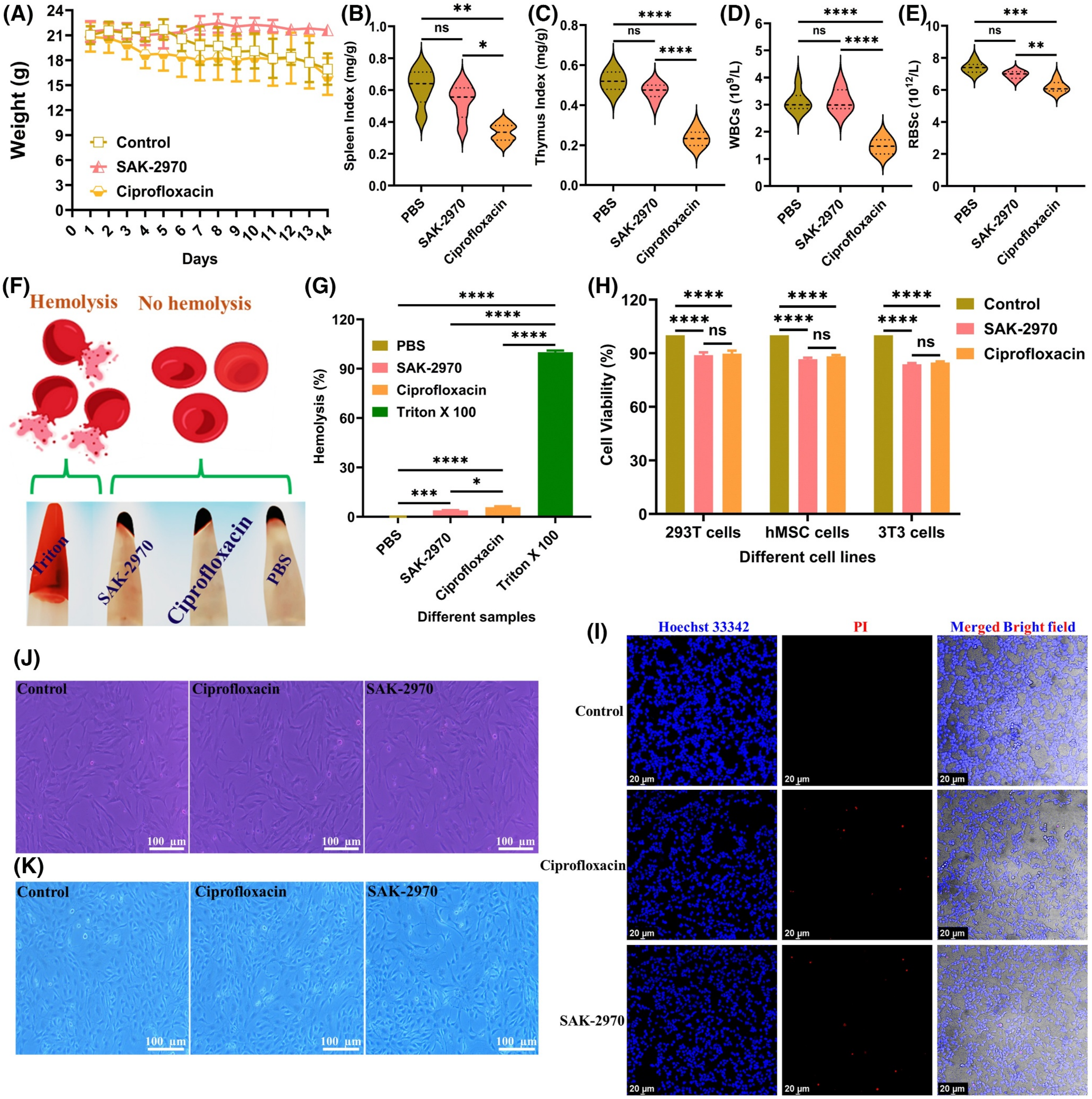

FIGURE 10 Safety and biocompatibility evaluation of SAK-2970 compared to ciprofloxacin. (A) Acute toxicity evaluation, depicted by mouse body weight changes after administration of 1500 mg/kg dose of SAK-2970 and ciprofloxacin ($n = 6$) for 14 days, with statistical significance indicated: ***$p = 0.0001$ (control vs. SAK-2970); *$p = 0.0328$ (control vs. ciprofloxacin); ****$p < 0.0001$ (SAK-2970 vs. ciprofloxacin). (B) Spleen index. Spleen index (mg/g) = spleen weight (mg)/body weight of the mice (G). Statistical analysis: ns = $p > 0.280$ (PBS vs. SAK-2970); **$p = 0.001$ (PBS vs. ciprofloxacin); *$p = 0.015$ (SAK-2970 vs. ciprofloxacin) ($n = 6$). (C) Thymus index. Thymus index (mg/g) = thymus weight (mg)/body weight of the mice (G). Statistics: ns = $p > 0.19$ (PBS vs. SAK-2970); ****$p < 0.0001$ (PBS vs. ciprofloxacin and SAK-2970 vs. ciprofloxacin) ($n = 6$). (D) Count of white blood cells (WBCs). Statistical analysis: ns = $p > 0.980$ (PBS vs. SAK-2970); ****$p < 0.0001$ (PBS vs. ciprofloxacin and SAK-2970 vs. ciprofloxacin) ($n = 6$). (E) Count of red blood cells (RBCs). Statistics: ns = $p > 0.08$ (PBS vs. SAK-2970); ***$p = 0.0002$ (PBS vs. ciprofloxacin); **$p = 0.004$ (SAK-2970 vs. ciprofloxacin) ($n = 6$). (F) Hemolysis assay depiction, distinguishing hemolytic activity with visual clarity. (G) Quantitative hemolysis analysis, presenting the hemolytic percentages of SAK-2970 and ciprofloxacin relative to PBS (negative control) and Triton X-100 (positive control), underscoring the superior hemocompatibility of SAK-2970 (*$p = 0.026$, ***$p = 0.0003$, and ****$p < 0.0001$). (H) Cytocompatibility assessment of SAK-2970 on normal cell lines (293T, hMSCs, and 3T3), with cell viability (%) comparable to ciprofloxacin, confirming high cellular compatibility (ns = $p > 0.05$, and ****$p < 0.0001$). (I) CLSM images of 293T cells treated with SAK-2970 and ciprofloxacin, revealing preserved nuclear integrity and membrane stability. (J and K) Optical microscopy images of hMSCs and 3T3 cells, respectively, following treatment, illustrating consistency of cell morphology with the control, indicative of SAK-2970's favorable cytocompatibility. Data are expressed as mean ± SD of three independent experiments.

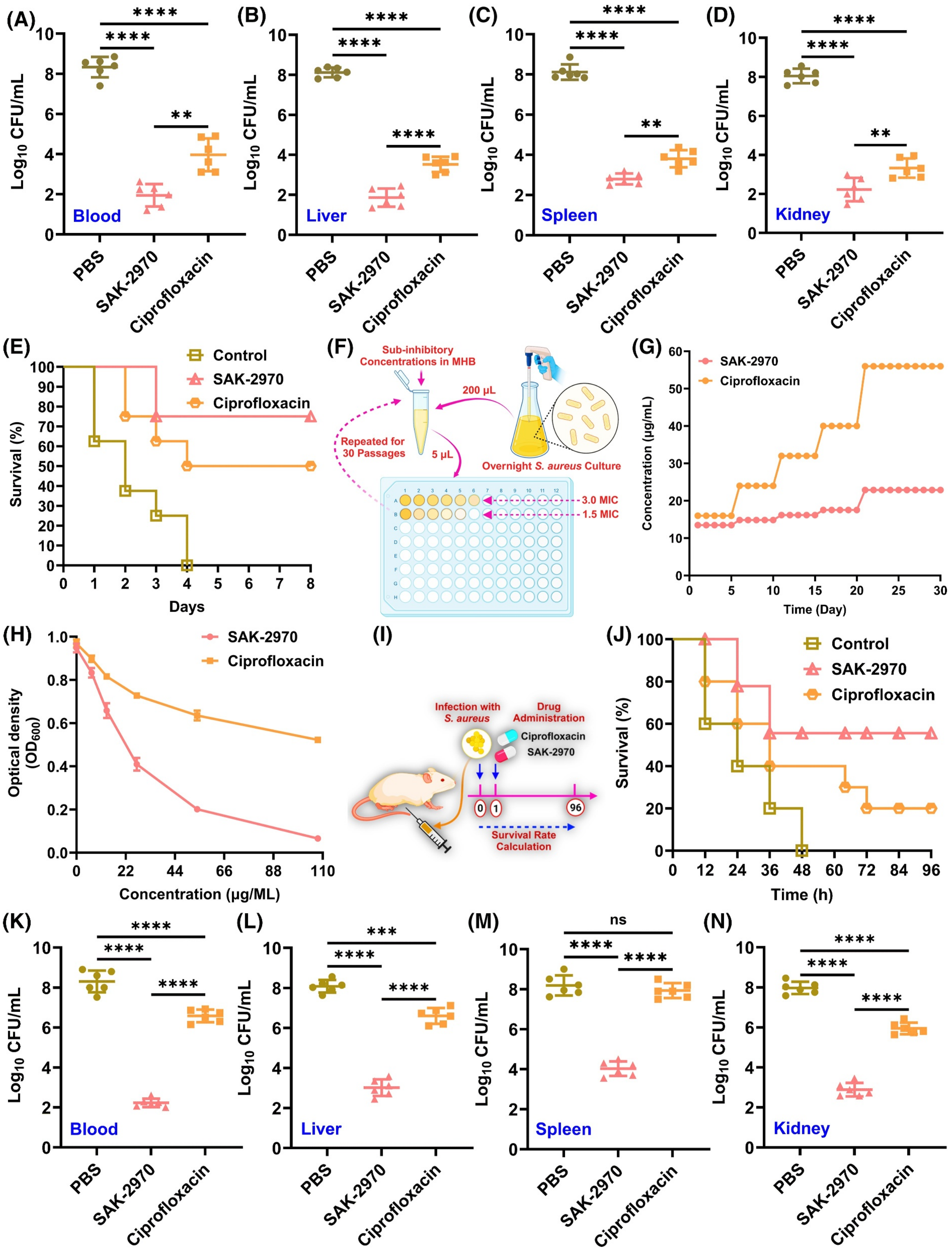

FIGURE 11 Legend on next page.

and lungs (Supplementary Figure S10B) of SAK-2970-treated mice. This substantial bacterial clearance was validated by statistically significant differences, attesting the SAK-2970's robust efficacy in controlling systemic infections. Additionally, as demonstrated in the survival analysis (Figure 11E), SAK-2970 administration yielded a notable survival benefit, with survival rates >70% compared to 50% with ciprofloxacin. Taken together, these results highlight the potential of SAK-2970 as a potent and clinically applicable agent for managing severe *S. aureus*-induced infections, presenting an innovative strategy to combat antimicrobial resistance and improve therapeutic outcomes in sepsis treatment.

3.13 | In vitro/in vivo efficacy of SAK-2970 against resistant *S. aureus*

Antimicrobial resistance remains a pivotal concern affecting the clinical deployment of antimicrobial agents. To evaluate the potential for the development of resistance against SAK-2970, an in vitro study was carried out in which *S. aureus* was continuously exposed to sub-inhibitory concentrations of SAK-2970 over a span of 30 consecutive days (Figure 11F). Remarkably, the propensity for resistance emergence was significantly lower for SAK-2970 than for the positive control, ciprofloxacin (Figure 11G). Moreover, a similar rapid escalation in ciprofloxacin resistance in *S. aureus* has also been reported previously.[58] Subsequently, the inhibitory efficacy of SAK-2970 was assessed against *S. aureus* strains that had developed resistance to ciprofloxacin. The findings revealed that SAK-2970 exerted a substantially enhanced inhibitory effect on ciprofloxacin-resistant *S. aureus* (Figure 11H). These results collectively suggest that SAK-2970 offers superior advantages in circumventing bacterial resistance mechanisms, highlighting its potential as a robust antimicrobial candidate with a diminished likelihood of resistance development.

To rigorously assess the in vivo antibacterial efficacy of SAK-2970 against resistant strains, we developed a mouse sepsis model infected with ciprofloxacin-resistant *S. aureus* (Figure 11I). Remarkably, survival analysis revealed that the survival rate of mice treated with 60 mg/kg SAK-2970 reached ~60% within 96 h post-infection, whereas the ciprofloxacin-treated group demonstrated a reduced survival rate of 20% under equivalent conditions (Figure 11J). These findings underscore the significant in vivo efficacy of SAK-2970 in combating drug-resistant bacteria, thereby offering a notable survival advantage. Further corroborating these results, bacterial load quantification 24 h post-SAK-2970 treatment revealed a marked reduction in CFUs across the liver, spleen, kidneys, and blood (Figure 11K–N), as well as the lungs (Supplementary Figure S10C), thereby affirming the compound's potent inhibitory action against ciprofloxacin-resistant *S. aureus* invasion. Collectively, these findings suggest that SAK-2970 holds promising therapeutic potential in mitigating the severity of infections caused by ciprofloxacin-resistant *S. aureus*.

4 | DISCUSSION

This study introduces a significant advancement in generative antibiotic discovery through the development of TargetGen-RNN, a GDL model designed to generate synthetically accessible, structurally novel, and biologically active small molecules under data-limited conditions. In contrast to distribution-agnostic or rigidly goal-conditioned generative paradigms, TargetGen-RNN integrates probabilistic latent space learning with temperature-scaled sampling[59] and minimal target-guided fine-tuning. The temperature control mechanism enables

FIGURE 11 (A–D) Bacterial load quantification in the blood, liver, spleen, and kidneys of *Staphylococcus aureus* bacteremia-infected mouse model post-treatment, showing superior clearance efficacy of SAK-2970 over ciprofloxacin, with significance levels: $^{**}p = 0.0013$, $^{****}p < 0.0001$ (blood); $^{****}p < 0.0001$ (liver); $^{**}p = 0.0025$, $^{****}p < 0.0001$ (spleen) and $^{**}p = 0.0060$, $^{****}p < 0.0001$ (kidneys) ($n = 6$). (E) Survival curves of *the S. aureus* bacteremia-infected mouse model ($n = 8$) treated with SAK-2970 and ciprofloxacin via oral gavage, demonstrating enhanced survival with SAK-2970 treatment. (F) Schematic representation of the *S. aureus* resistance development assay under sequential subculturing with sub-inhibitory doses of SAK-2970 and ciprofloxacin. Created with BioRender.com (https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en). (G) Comparative assessment of resistance progression in *S. aureus* exposed to SAK-2970 versus ciprofloxacin over 30 days, indicating slower resistance development with SAK-2970. (H) Dose-dependent inhibition of ciprofloxacin-resistant *S. aureus* by SAK-2970, underscoring its potent efficacy at the tested concentrations. (I) Schematic representation of the study model for anti-*S. aureus*, illustrating the infection and treatment protocols. Created with BioRender.com (https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en). (J) Survival curves of mouse model infected with ciprofloxacin-resistant *S. aureus* ($n = 10$) and treated via oral gavage with SAK-2970 or ciprofloxacin, showing a pronounced survival benefit with SAK-2970. (K–N) Bacterial burden quantification in blood, liver, spleen, and kidneys in the ciprofloxacin-resistant *S. aureus* bacteremia mouse model, highlighting SAK-2970's superior efficacy in bacterial clearance, with significance levels: $^{****}p < 0.0001$ (blood); $^{***}p = 0.0001$, $^{****}p < 0.0001$ (liver); ns = $p > 0.50$, $^{****}p < 0.0001$ (spleen), and $^{****}p < 0.0001$ (kidneys) ($n = 6$).

dynamic modulation of generative entropy, facilitating broad exploration of chemical space at elevated temperatures and structurally focused outputs under more constrained conditions.[59,60] This adaptive sampling strategy surpasses static methods employed by conventional generative architectures, including VAEs,[21] GANs,[22] and cRNNs,[29] thereby enabling simultaneous scaffold innovation and pharmacophoric specificity. To ensure the reliability and quality of generated compounds, a two-tier validation framework was implemented. Chemical correctness was verified using RDKit, which enforced standardized valency and bonding constraints across all SMILES representations. Structural novelty was quantified using Tanimoto similarity analysis relative to both the pretraining and fine-tuning datasets. This dual-layer validation confirmed that the 28,708 molecules generated by TargetGen-RNN were not only chemically viable but also structurally distinct, populating underexplored regions of chemical space beyond the model's prior exposure.

Performance benchmarking further demonstrated the superiority of TargetGen-RNN over a reference cRNN model,[29] achieving a higher final predictive accuracy (~0.90 vs. ~0.85) and reduced convergence loss (~0.30–0.40 vs. ~0.45) (Supplementary Figure S1). To further contextualize the performance, the model was benchmarked against additional state-of-the-art GLD models using established evaluation metrics. As summarized in Table 1, TargetGen-RNN consistently outperformed the comparators, generating chemically valid, unique, and novel molecules with enhanced scaffold diversity and pharmacophoric relevance. Notably, the model yielded a novelty score of 0.999, filter pass rate of 99.8%, and high uniqueness across sample sizes, indicating low output redundancy. Internal diversity scores (InDiv1: 0.891; InDiv2: 0.889) reflected broad yet chemically coherent sampling of the latent space. Although the direct retraining of all comparator models under identical conditions would provide strict head-to-head validation, the benchmarking approach employed here is consistent with accepted practices and offers a robust comparative framework.

Scaffold-level analyses further supported the extrapolative capacity of the model. The Bemis–Murcko enumeration and Tanimoto distance metrics revealed a marked topological divergence between the training and fine-tuning data. This scaffold innovation emerged not from random dispersion but from a chemically structured pharmacophore-aware generative process. The resulting compounds exhibited favorable drug-like properties, elevated QED scores, acceptable synthetic accessibility, and a low incidence of PAINS or structural alerts. Quantitatively, TargetGen-RNN achieved a scaffold diversity of 83.81% and a target-novelty score of 99.99%, outperforming the benchmarks reported for the existing GDL model.[29] Collectively, these findings establish TargetGen-RNN as a high-fidelity, generalizable platform for scaffold innovation and biologically relevant compound generation in sparse data environments. The model's integrated architecture, rigorous validation pipeline, and favorable performance metrics underscore its translational potential for de novo drug discovery.

From the generative ensemble, SAK-2970 emerged as a lead scaffold with compelling biophysical and pharmacological attributes. MD simulations revealed that SAK-2970 adopts a dynamically stable and pharmacophorically optimized binding mode to DHFR, marked by minimal RMSD fluctuations, low atomic flexibility, and persistent hydrogen bonding and π–π stacking. This interaction profile closely mirrors that of the reference ligand, suggesting that TargetGen-RNN effectively captures and reproduces high-affinity binding patterns, even when trained on limited target-specific data. The observed rigidity and interaction density indicate a sterically and

TABLE 1 Comparative performance metrics of reported models (AAE,[61] CharRNN,[25] VAE,[21,62] ORGAN,[63] DNMG,[64] cRNN[29]) and TargetGen-RNN (our model) across multiple sample sizes of the generated molecules.

	Comparative performance metrics of models for molecule generation								
Different models	**Valid**	**Unique@1k**	**Unique@10k**	**Unique@total**	**Novelty**	**Filters**	**InDiv1 (1k)**	**InDiv2 (10k)**	**InDiv3 (total)**
AAE	0.957	1	0.997	NA	0.918	0.994	0.853	0.85	NA
CharRNN	0.996	1	0.996	NA	0.773	0.995	0.855	0.841	NA
VAE	0.941	1	0.998	NA	0.897	0.992	0.852	0.842	NA
ORGAN	0.734	0.987	0.93	NA	1	0.995	0.66	0.644	NA
DNMG	0.999	1	0.998	NA	0.936	0.996	0.856	0.85	NA
cRNN	1	1	1	NA	0.997	NA	NA	NA	NA
TargetGen-RNN (our model)	1	1	1	1	0.999	0.998	0.891	0.889	0.889

Abbreviations: cRNN, conditional RNN; RNN, recurrent neural network; VAE, variational autoencoder.

energetically favorable fit within the DHFR active site, reinforcing the suitability of SAK-2970 as a lead scaffold. In contrast, compounds such as SAK-1893 and SAK-3702 exhibited elevated ligand RMSDs and fewer stabilizing contacts, which is consistent with suboptimal orientation or transient binding. These differences likely reflect scaffold-specific conformational dynamics rather than instability, underscoring the limitations of static docking and the importance of dynamic validation. Compared to prior MD studies on *S. aureus* targets, such as the YsxC–ligand complexes reported by Kumari et al., which showed protein and ligand RMSDs of 2.9–3.7 Å and up to 1.0 Å, respectively,[65] the SAK-2970–DHFR complex maintained tighter conformational control. These findings highlight the critical role of long-timescale MD simulations in refining generative design outputs and prioritizing candidates with sustained binding and translational relevance.[66,67]

Moreover, the MM/GBSA binding free energy analysis highlighted the dominant van der Waals (ΔG_{Bind}vdW) and Coulombic (ΔG_{Bind}Coulomb) contributions, reflecting a favorable enthalpic profile (Supplementary Table S4). Lipophilic interactions (ΔG_{Bind}Lipo) also supported complex stabilization. In contrast, desolvation penalties (ΔG_{Bind}SolvGB) exhibited positive values, partially offsetting the net binding affinity, although they remained within energetically compensable thresholds. Covalent interactions (ΔG_{Bind}Covalent) were minimal, with slightly positive or near-neutral contributions across most complexes. Collectively, these findings indicate that DHFR–ligand complex stability is primarily driven by noncovalent interactions, with van der Waals and electrostatic forces serving as the predominant forces.

Empirical microbiological assays substantiated the in silico findings. SAK-2970 exhibited potent bactericidal activity against both drug-sensitive and ciprofloxacin-resistant *S. aureus* strains, with sub-micromolar MIC and MBC values and rapid concentration-dependent killing kinetics. SAK-2970 demonstrated potent DHFR inhibition ($IC_{50} = 2.379$ nM), substantially exceeding the activity of previously reported leads such as compound CTh7 ($IC_{50} = 150$ nM).[68] In contrast, compound 6 achieved the highest DHFR inhibition in its study ($IC_{50} = 3.731$ μg/mL), surpassing trimethoprim's effectiveness by approximately threefold.[69] Given that SAK-2970 displayed greater potency than compound 6, it is also inferred to outperform trimethoprim in enzymatic inhibition. SAK-2970 further demonstrated pronounced antibiofilm efficacy, achieving >90% inhibition at sub-MIC levels and >85% biofilm eradication at 2 × MIC concentrations. Notably, its efficacy extended to biofilm-embedded and resistance-induced phenotypes, suggesting a mechanism of action that is resilient to conventional resistance pathways. Safety and pharmacokinetic assessments further reinforced the translational viability of this compound. SAK-2970 exhibited negligible hemolytic activity and minimal cytotoxicity in mammalian cell lines, whereas in silico ADMET profiling indicated high oral bioavailability, metabolic stability, and absence of hERG liability. In vivo validation using murine systemic infection models demonstrated significant bacterial clearance across primary organs and superior survival rates relative to ciprofloxacin in both wild-type and resistant bacterial infection cohorts.

Although this study establishes the efficacy of TargetGen-RNN in generating DHFR-targeting antibacterial agents, several strategic directions remain to fully realize its translational scope. The deliberate focus on a single, clinically validated enzyme enabled a rigorous assessment of the model's generative precision; however, TargetGen-RNN application to a broader repertoire of bacterial proteins, particularly those exhibiting diverse topologies or resistance phenotypes, will be essential to assess its scalability across distinct ligand-binding landscapes. While acute in vivo models have confirmed both therapeutic activity and safety, rigorous pharmacokinetic and toxicological characterization in chronic and immunocompromised settings remains critical for advancing clinical translation. Importantly, the modular architecture of TargetGen-RNN facilitates its generalization to other clinically relevant bacterial targets, including gram-negative pathogens such as *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*. This adaptation requires domain-specific retraining/fine-tuning using appropriate ligand datasets. Given their underrepresentation in current antimicrobial pipelines, these extensions offer high translational value. Beyond infectious diseases, domain-specific fine-tuning, such as with kinase inhibitors or modulators of protein–protein interactions, could enable the rational design of novel chemotypes in oncology and other complex indications. Collectively, these trajectories underscore the potential of structure-informed generative modeling to accelerate lead discovery across mechanistically diverse and clinically urgent target classes.

5 | CONCLUSION

In conclusion, this study unequivocally demonstrated the potential of TargetGen-RNN as a transformative tool for discovering novel antibiotics targeting drug-resistant *S. aureus*. By employing advanced deep generative learning with target-specific fine-tuning, the model efficiently navigates the chemical space to yield novel, diverse, and biologically relevant compounds. In-depth analysis and screening of the generated compounds revealed candidates with significant antibacterial efficacy, particularly SAK-2970, which emerged as a superlative candidate, validated by its formidable antibacterial and anti-biofilm efficacies in vitro, including bactericidal activity against ciprofloxacin-resistant strains. In vivo analyses further substantiated SAK-2970's therapeutic potential, evidencing a significant reduction in bacterial burden and increased survival rates in a mouse model of drug-resistant bacteremia, with minimal systemic toxicity, highlighting its biocompatibility and therapeutic potential. The success of TargetGen-RNN in producing structurally novel molecules aligned with desired pharmacokinetic profiles validates its value in overcoming the limitations associated with traditional antibiotic discovery. These findings not only suggest SAK-2970 as a viable candidate for combating antimicrobial resistance but also exemplify a transformative paradigm in de novo antibiotic design. Future research endeavors may further amplify the applicability of TargetGen-RNN across disparate pathogens and resistance mechanisms, thereby propelling its utility in the development of next-generation therapeutics.

ACKNOWLEDGMENTS

This publication is based upon work supported by King Fahd University of Petroleum & Minerals. Author(s) at KFUPM acknowledge the Interdisciplinary Research Center for Biosystems and machines for the support received under Grant No. INBS2501. The authors wish to extend their heartfelt appreciation to the Department of Applied Biology and Chemical Technology at the Hong Kong Polytechnic University in Hong Kong, China.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ETHICS STATEMENT

All animal experiments were performed under the protocols approved by the Institutional Biosafety and Bioethics Committee (IBC) (Reference no. 5010/ORIC), University of Agriculture Faisalabad (UAF), Faisalabad-Pakistan.

ORCID

Shakeel Ahmad Khan https://orcid.org/0000-0003-0967-3079
Adnan Shakoor https://orcid.org/0000-0003-0646-5231

REFERENCES

1. S. Y. C. Tong, J. S. Davis, E. Eichenberger, T. L. Holland, V. G. Fowler, *Clin. Microbiol. Rev.* **2015**, *28*, 603.
2. G. Y. Fang, F. H. Wu, X. J. Mu, Y. J. Jiang, X. Q. Liu, *J. Hazard. Mater.* **2024**, *465*, 133136.
3. Y. Chen, Z. Liu, Z. Lin, M. Lu, Y. Fu, G. Liu, B. Yu, *Front. Immunol.* **2023**, *14*, 1219895.
4. Q. H. Yu, R. Huang, K. Y. Wu, X. L. Han, Y. J. Cheng, W. L. Liu, A. Q. Zhang, S. Y. Qin, *Acta Biomater.* **2022**, *154*, 359.
5. R. Huang, Q. H. Yu, X. D. Yao, W. L. Liu, Y. J. Cheng, Y. H. Ma, A. Q. Zhang, S. Y. Qin, *ACS Appl. Mater. Interfaces* **2022**, *14*, 159.
6. D. Y. Zhang, R. G. Cao, Y. J. Cheng, W. L. Liu, R. Huang, A. Q. Zhang, S. Y. Qin, *J. Control. Release* **2023**, *362*, 565.
7. B. P. Howden, S. G. Giulieri, T. Wong Fok Lung, S. L. Baines, L. K. Sharkey, J. Y. H. Lee, A. Hachani, I. R. Monk, T. P. Stinear, *Nat. Rev. Microbiol.* **2023**, *21*, 380.
8. S. Bishen, What I've Learnt from a Decade of Campaigning on Superbugs - Dame Sally Davies, https://www.weforum.org/agenda/2024/08/antimicrobial-resistance-superbugs-antibiotics/ (accessed: August, 2024).
9. Y. Guo, G. Song, M. Sun, J. Wang, Y. Wang, *Front. Cell. Infect. Microbiol.* **2020**, *10*, 511382.
10. P. S. Loomba, J. Taneja, B. M. Mishra, *J. Global Infect. Dis.* **2010**, *2*, 275.
11. F. F. Tuon, P. H. Suss, J. P. Telles, L. R. Dantas, N. H. Borges, V. S. T. Ribeiro, *Antibiotics* **2023**, *12*, 87.
12. J. A. Lindsay, *Int. J. Med. Microbiol.* **2010**, *300*, 98.
13. X. Zhu, Q. Tang, X. Zhou, M. R. Momeni, *Microb. Pathog.* **2024**, *193*, 106741.
14. Y. A. Ivanenkov, R. S. Yamidanov, I. A. Osterman, P. V. Sergiev, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, M. S. Veselov, A. A. Ayginin, D. A. Skvortsov, K. S. Komarova, S. V. Sadovnikov, R. Matniyazov, A. A. Sofronova, A. S. Malyshev, A. E. Machulkin, R. A. Petrov, D. Lukianov, S. Iarovenko, D. S. Bezrukov, A. K. Baymiev, O. A. Dontsova, *J. Antibiot.* **2019**, *72*, 827.
15. R. Cain, S. Narramore, M. McPhillie, K. Simmons, C. W. G. Fishwick, *Bioorg. Chem.* **2014**, *55*, 69.
16. N. S. Togre, A. M. Vargas, G. Bhargavi, M. K. Mallakuntla, S. Tiwari, *Int. J. Mol. Sci.* **2022**, *23*, 10669.
17. A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, I. E. Orhan, M. Banach, J. M. Rollinger, D. Barreca, W. Weckwerth, R. Bauer, E. A. Bayer, M. Majeed, A. Bishayee, V. Bochkov, G. K. Bonn, N. Braidy, F. Bucar, A. Cifuentes, G. D'Onofrio, M. Bodkin, M. Diederich, A. T. Dinkova-Kostova, T. Efferth, K. El Bairi, N. Arkells, T. P. Fan, B. L. Fiebich, M. Freissmuth, M. I. Georgiev, S. Gibbons, K. M. Godfrey, C. W. Gruber, J. Heer, L. A. Huber, E. Ibanez, A. Kijjoa, A. K. Kiss, A. Lu, F. A. Macias, M. J. S. Miller, A. Mocan, R. Müller, F. Nicoletti, G. Perry, V. Pittalà, L. Rastrelli, M. Ristow, G. L. Russo, A. S. Silva, D. Schuster, H. Sheridan, K. Skalicka-Woźniak, L. Skaltsounis, E. Sobarzo-Sánchez, D. S. Bredt, H. Stuppner, A. Sureda, N. T. Tzvetkov, R. A. Vacca, B. B. Aggarwal, M. Battino, F. Giampieri, M. Wink, J. L. Wolfender, J. Xiao, A. W. K. Yeung, G. Lizard, M. A. Popp, M. Heinrich, I. Berindan-Neagoe, M. Stadler, M. Daglia, R. Verpoorte, C. T. Supuran, *Nat. Rev. Drug Discov.* **2021**, *20*, 200.
18. X. Yang, Y. Wang, R. Byrne, G. Schneider, S. Yang, *Chem. Rev.* **2019**, *119*, 10520.
19. T. Sousa, J. Correia, V. Pereira, M. Rocha, *J. Chem. Inf. Model.* **2021**, *61*, 5343.
20. X. Zeng, F. Wang, Y. Luo, S. G. Kang, J. Tang, F. C. Lightstone, E. F. Fang, W. Cornell, R. Nussinov, F. Cheng, *Cell Rep. Med.* **2022**, *3*, 100794.
21. R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, *ACS Cent. Sci.* **2018**, *4*, 268.
22. A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A.

Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, *Nat. Biotechnol.* **2019**, *37*, 1038.
23. Y. Li, J. Pei, L. Lai, *Chem. Sci.* **2021**, *12*, 13664.
24. J. Wang, C. Y. Hsieh, M. Wang, X. Wang, Z. Wu, D. Jiang, B. Liao, X. Zhang, B. Yang, Q. He, D. Cao, X. Chen, T. Hou, *Nat. Mach. Intell.* **2021**, *3*, 914.
25. M. H. S. Segler, T. Kogej, C. Tyrchan, M. P. Waller, *ACS Cent. Sci.* **2018**, *4*, 120.
26. N. Brown, M. Fiscato, M. H. S. Segler, A. C. Vaucher, *J. Chem. Inf. Model.* **2019**, *59*, 1096.
27. P. Renz, D. Van Rompaey, J. K. Wegner, S. Hochreiter, G. Klambauer, *Drug Discov. Today Technol.* **2019**, *32–33*, 55.
28. J. Meyers, B. Fabian, N. Brown, *Drug Discov. Today* **2021**, *26*, 2707.
29. Y. Li, L. Zhang, Y. Wang, J. Zou, R. Yang, X. Luo, C. Wu, W. Yang, C. Tian, H. Xu, F. Wang, X. Yang, L. Li, S. Yang, *Nat. Commun.* **2022**, *13*, 6891.
30. M. Moret, L. Friedrich, F. Grisoni, D. Merk, G. Schneider, *Nat. Mach. Intell.* **2020**, *2*, 171.
31. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman, *J. Chem. Inf. Model.* **2012**, *52*, 1757.
32. M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, *Nucleic Acids Res.* **2016**, *44*, D1045.
33. N. P. Muddala, J. C. White, B. Nammalwar, I. Pratt, L. M. Thomas, R. A. Bunce, K. D. Berlin, C. R. Bourne, *Eur. J. Med. Chem.* **2020**, *200*, 112412.
34. M. Kobayashi, T. Kinjo, Y. Koseki, C. R. Bourne, W. W. Barrow, S. Aoki, *J. Chem. Inf. Model.* **2014**, *54*, 1242.
35. C. A. Lipinski, *Drug Discov. Today Technol.* **2004**, *1*, 337.
36. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615.
37. A. K. Ghose, V. N. Viswanadhan, J. J. A. Wendoloski, *J. Comb. Chem.* **1999**, *1*, 55.
38. S. L. Dixon, J. Duan, E. Smith, C. D. Von Bargen, W. Sherman, M. P. Repasky, *Future Med. Chem.* **2016**, *8*, 1825.
39. RCSB PDB – 4LAE: Structure-Based Design of New Dihydrofolate Reductase Antibacterial Agents: 7-(Benzimidazol-1-Yl)-2,4-Diaminoquinazolines, https://www.rcsb.org/structure/4LAE (accessed: August 2024).
40. *Schrödinger Release 2023-1. Glide*, Schrödinger, LLC, New York, NY **2023**.
41. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, *J. Med. Chem.* **2006**, *49*, 6177.
42. Schrödinger Release 2023-1, *Protein Preparation Wizard*, Epik, Schrödinger, LLC, New York, NY **2023**.
43. M. Burgdorf, A. Robinson-Mosher, X. Guo, K. H. Law, D. E. Shaw, Desmond/GPU performance as of April 2021, https://www.deshawresearch.com/publications/Desmond-GPU_Performance_April_2021.pdf (accessed: March, 2021).
44. S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, *9*, 1735.
45. L. McInnes, J. Healy, N. Saul, L. Großberger, *J. Open Source Softw.* **2018**, *3*, 861.
46. F. Grisoni, B. J. H. Huisman, A. L. Button, M. Moret, K. Atz, D. Merk, G. Schneider, *Sci. Adv.* **2021**, *7*, eabg3338.
47. G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887.
48. M. González-Medina, J. L. Medina-Franco, *J. Chem. Inf. Model.* **2017**, *57*, 1735.
49. R. Brenk, A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson, P. G. Wyatt, *ChemMedChem* **2008**, *3*, 435.
50. E. F. Marondedze, K. K. Govender, P. P. Govender, *J. Mol. Graphics Modell.* **2020**, *101*, 107711.
51. M. El-Naggar, H. A. Sallam, S. S. Shaban, S. S. Abdel-Wahab, A. E. G. E. Amr, M. E. Azab, E. S. Nossier, M. A. Al-Omar, *Molecules* **2019**, *24*, 1066.
52. I. M. M. Othman, M. A. M. Gad-Elkareem, A. E. G. E. Amr, M. A. Al-Omar, E. S. Nossier, E. A. Elsayed, *J. Enzyme Inhib. Med. Chem.* **2020**, *35*, 1491.
53. S. Roy, M. Halder, P. Ramprasad, S. Dasgupta, Y. Singh, D. Pal, *Int. J. Biol. Macromol.* **2023**, *249*, 126049.
54. R. Gaglione, A. Cesaro, E. Dell'olmo, R. Di Girolamo, L. Tartaglione, E. Pizzo, A. Arciello, *Int. J. Mol. Sci.* **2020**, *21*, 2049.
55. R. Kant, V. Singh, G. Nath, S. K. Awasthi, A. Agarwal, *Eur. J. Med. Chem.* **2016**, *124*, 218.
56. K. Amin, R. M. Dannenfelser, *J. Pharm. Sci.* **2006**, *95*, 1173.
57. D. G. Pyun, H. S. Yoon, H. Y. Chung, H. J. Choi, T. Thambi, B. S. Kim, D. S. Lee, *J. Mater. Chem. B* **2015**, *3*, 7752.
58. H. M. Blumberg, D. Rimland, D. J. Carroll, P. Terry, I. Kaye Wachsmuth, *J. Infect. Dis.* **1991**, *163*, 1279.
59. J. A. Weller, R. Rohs, *J. Chem. Inf. Model.* **2024**, *64*, 6450.
60. R. Child, in *ICLR 2021 – 9th International Conference on Learning Representations*, Vienna, Austria **2021**.
61. B. Ghojogh, M. Crowley, F. Karray, A. Ghodsi, *Elements of dimensionality reduction and manifold learning*, Springer, Cham, Switzerland **2023**, Ch. 21.
62. D. P. Kingma, M. Welling, *(Preprint) arXiv*, 1312.6114v11, **2022**.
63. G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias, A. Aspuru-Guzik, *(Preprint) arXiv*, 1705.10843v3, **2018**.
64. T. Song, Y. Ren, S. Wang, P. Han, L. Wang, X. Li, A. Rodriguez-Patón, *Methods* **2023**, *211*, 10.
65. R. Kumari, R. Rathi, S. R. Pathak, V. Dalal, *J. Mol. Struct.* **2022**, *1255*, 132476.
66. R. Kumari, V. Dalal, *J. Biomol. Struct. Dyn.* **2022**, *40*, 9833.
67. V. Dalal, P. Dhankhar, V. Singh, V. Singh, G. Rakhaminov, D. Golemi-Kotra, P. Kumar, *Protein J.* **2021**, *40*, 148.
68. R. Sehrawat, R. Pasrija, P. Rathee, D. Kumari, A. Khatkar, E. Küpeli Akkol, E. Sobarzo-Sánchez, *Antibiotics* **2024**, *13*, 479.
69. L. M. A. A. Ghany, N. Ryad, M. S. Abdel-Aziz, H. M. A. El-Lateef, I. Zaki, B. Y. Beshay, *J. Mol. Struct.* **2024**, *1309*, 138170.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: S. A. Khan, A. Shakoor, S. Kanwal, *Interdiscip. Med.* **2026**, *4*, e70064. https://doi.org/10.1002/inmd.70064