applied biosciences

MDPI

Article

An Exploratory Reinforcement Learning Simulation Framework for Studying Antimicrobial Resistance Dynamics Under Copper Exposure

Hayden D. Hedman

School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332, USA; hhedman3@gatech.edu

Abstract

This study presents an exploratory reinforcement learning (RL)-based simulation framework for examining antimicrobial resistance (AMR) dynamics under repeated exposure to a non-antibiotic stressor, using copper as a simplified model compound. The objective is not to provide mechanistic or predictive insight into microbial evolution, but to evaluate how alternative sequential decision-making strategies perform within a constrained and transparent simulation environment. Three agent strategies were compared: random action selection, a rule-based heuristic, and a tabular Q-learning agent. Simulations were conducted over fixed 40-cycle episodes in which agents adjusted copper exposure in response to evolving resistance-related state variables. Across experimental runs, the Q-learning agent exhibited lower cumulative resistance burden, measured by area under the curve (AUC) of minimum inhibitory concentration (MIC) trajectories for chloramphenicol and polymyxin B, while maintaining lower cumulative copper exposure relative to baseline strategies. The rule-based agent demonstrated intermediate performance, whereas the random agent showed greater variability and less stable trajectories. These findings reflect differences in simulated control behavior within a simplified stochastic system. Overall, this work introduces an interpretable reinforcement learning simulation tool intended to support comparative evaluation of adaptive versus static strategies in antimicrobial pressure management under limited observability.

Keywords: reinforcement learning; antimicrobial resistance; copper exposure; minimum inhibitory concentration (MIC); Q-learning; agent-based modeling; simulation framework

Check for updates

Academic Editor: Viviana Maresca

Received: 21 January 2026
Revised: 6 March 2026
Accepted: 20 April 2026
Published: 3 May 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Background on AMR and Metal Stressors

Antimicrobial resistance (AMR) continues to be one of the most critical global health challenges, with resistance to conventional antibiotics increasing at an alarming rate worldwide [1–3]. The widespread emergence of resistant pathogens has undermined the long-term effectiveness of many frontline antimicrobial agents, placing increasing strain on clinical, agricultural, and public health systems [4–6]. While antibiotic development remains an important priority, the pace of novel drug discovery has not kept up with the rate at which resistance evolves [7,8], motivating interest in complementary strategies for managing antimicrobial pressure rather than relying exclusively on new synthetic compounds [9–11].

Within this context, alternative antimicrobial agents have attracted growing attention as potential tools for modulating resistance dynamics [12–14]. Among these, copper

has long been recognized for its broad-spectrum antimicrobial activity and its capacity to inhibit microbial growth through multiple, non-specific mechanisms [15–18]. Copper and copper-based materials have been explored across diverse settings, including surface coatings, agricultural applications, and environmental controls, where they may reduce microbial burden without acting as direct therapeutic replacements for antibiotics [15,16,18]. However, while copper shows consistent antimicrobial activity, evidence on its longer-term influence on AMR trajectories is mixed and remains incompletely characterized across settings and exposure regimes [12,19,20]. In particular, the interaction between sublethal copper exposure and the evolution of antibiotic resistance poses unresolved questions, especially given evidence that metal stressors can contribute to co-selection and persistence of resistance phenotypes under certain conditions [17,21]. The framework introduced here does not attempt to mechanistically replicate population genetics, mutation rates, or horizontal gene transfer processes. Instead, it provides a simplified stochastic environment designed to examine how alternative decision strategies behave under repeated antimicrobial pressure.

Rather than serving as a substitute for conventional antimicrobial therapy, copper provides a biologically grounded model compound for examining how selective pressure influences resistance dynamics over time. Its use as a model stressor allows investigation of resistance evolution under repeated, variable exposure without the confounding assumptions associated with optimized therapeutic dosing. This perspective motivates the need for systematic and adaptive experimentation frameworks capable of capturing how resistance accumulates, stabilizes, or partially reverses across sequential exposure cycles.

The evolution of antimicrobial resistance can be naturally framed as a sequential decision-making and feedback-control problem [22,23]. Resistance does not emerge from a single exposure event but instead reflects cumulative responses to repeated environmental pressures, where past decisions shape future susceptibility and growth outcomes. From this perspective, antimicrobial management involves a series of interdependent control decisions made under uncertainty, incomplete observability, and delayed feedback. Reinforcement learning (RL) provides a computational framework well suited to such settings, as it explicitly models sequential decision-making in dynamic environments where system dynamics may be stochastic or only partially known.

1.2. Related Work

Recent foundational work has demonstrated that reinforcement learning can learn adaptive intervention strategies capable of influencing resistance trajectories in complex evolutionary systems defined by antibiotic fitness landscapes, even under noisy or delayed feedback conditions [22,24]. These studies establish the feasibility of RL-based approaches for AMR control and illustrate how adaptive policies can outperform static treatment paradigms when resistance dynamics unfold over time. However, existing RL applications in AMR research have largely focused on synthetic antibiotics and highly structured treatment environments, often assuming detailed fitness landscapes, rich diagnostic information, or tightly controlled experimental conditions. However, existing RL applications in AMR research have largely focused on synthetic antibiotics within highly structured treatment environments, often assuming detailed fitness landscapes, rich diagnostic information, or tightly controlled experimental conditions. As a result, it remains unclear how adaptive learning strategies perform when applied to non-antibiotic stressors or within simplified, constrained environments where resistance must be inferred from coarse indicators rather than comprehensive evolutionary models. The present study addresses this gap by evaluating adaptive and heuristic control strategies within an intentionally simplified and transparent simulation framework designed to reflect such constraints.

As a result, comparatively little work has examined how adaptive decision-making frameworks behave when applied to non-antibiotic antimicrobial compounds or simplified settings that more closely reflect constrained clinical, agricultural, or infrastructural capacities. In many real-world environments, resistance management decisions must be made with limited diagnostic feedback, coarse measurements of susceptibility, and restricted ability to precisely modulate antimicrobial exposure. Understanding how learning-based and heuristic strategies perform under such constraints represents an important methodological gap.

1.3. Research Project Aims

Building on this foundation, the present study applies a reinforcement learning–based simulation framework using copper as a representative non-antibiotic antimicrobial stressor to explore how different decision-making strategies respond to antimicrobial pressure across variable and constrained environments. The goal of this work is not to evaluate the clinical effectiveness of copper itself, but to examine how varying levels of adaptability influence resistance dynamics under simplified, interpretable assumptions. By abstracting resistance evolution using commonly reported microbiological indicators such as minimum inhibitory concentrations (MICs) and derived resistance burden indices, the framework emphasizes conceptual clarity and methodological transparency rather than biological or computational complexity.

To this end, three agent strategies are considered: (1) a random agent representing uncontrolled or unstructured exposure, (2) a rule-based agent implementing fixed decision criteria based on observed resistance trends, and (3) a Q-learning agent employing adaptive decision-making through iterative interaction with the environment. By comparing these agents within a shared simulation environment, this study provides an exploratory framework for understanding how adaptive control strategies may perform relative to static or heuristic approaches when resistance evolves over repeated exposure cycles.

Collectively, this work situates antimicrobial resistance within a systems and control-oriented perspective, highlighting the potential role of reinforcement learning as a methodological tool for studying AMR dynamics under uncertainty. Rather than offering prescriptive treatment recommendations, the framework introduced here aims to support future investigations into AI-assisted modeling approaches that bridge microbiology, systems thinking, and adaptive decision-making in the study of antimicrobial resistance.

2. Materials and Methods

2.1. Overview of Study and Framework

This study implemented a reinforcement learning (RL)–based simulation framework to model antibiotic resistance dynamics under varying copper exposure conditions across discrete experimental cycles. The experimental setup was rooted in copper concentration estimates reported in the foundational study by Li et al. (2021) [21], with parameters adapted to simulate copper exposure and antibiotic resistance evolution in *Escherichia coli K12* and *Staphylococcus aureus*. The RL environment and agents were implemented in Python 3.11.9 (Gymnasium, NumPy, and Matplotlib) for simulation, computation, and figure generation. Simulations modeled bacterial populations exposed to varying copper concentrations in combination with antibiotics chloramphenicol and polymyxin B [21]. Three agents were evaluated: (1) a random policy agent, (2) a rule-based agent, and (3) a Q-learning agent. These models were designed to simulate control strategies for antimicrobial resistance dynamics under copper exposure.

Simulations were conducted over fixed-length episodes, each consisting of 40 cycles. At the start of each episode, the environment was reset to baseline conditions and the

agent state was reinitialized. Within each cycle, agents selected a copper adjustment action based on the current state, and resistance- and growth-related outcomes were updated according to stochastic transition rules. The transition rules governing resistance updates were intentionally heuristic and stochastic rather than mechanistic. They do not explicitly model mutation rates, horizontal gene transfer, fitness costs, or pharmacodynamic processes. Instead, the environment was designed to provide a transparent and computationally tractable control setting grounded in reported copper concentration ranges (21), allowing comparative evaluation of decision strategies under simplified and interpretable assumptions. Agent performance was evaluated by aggregating per-cycle outputs across repeated simulation runs. Simulations rely on pseudo-random number generation within the environment transition dynamics; repeated episodes were used to characterize variability in agent trajectories.

2.2. Reinforcement Learning Agent Design

The reinforcement learning (RL) agents interacted with a simulation environment designed to capture key features of copper-driven antimicrobial pressure and resistance dynamics over discrete experimental cycles [21]. The environment represents bacterial populations exposed to variable copper concentrations in the presence of antibiotics, without explicitly modeling underlying molecular or genetic mechanisms.

The environment state was represented at each cycle as a five-dimensional vector consisting of current copper concentration (mg/L), minimum inhibitory concentration (MIC) values for chloramphenicol and polymyxin B, cycle count, and a normalized resistance burden index. These state variables were selected to reflect commonly reported microbiological indicators of resistance and antimicrobial pressure, including MIC values and derived resistance burden summaries [25,26]. Copper concentration was bounded within an experimentally plausible range (0–100 mg/L), while MIC values were initialized at low baseline levels and allowed to evolve dynamically over time. The cycle count tracked progression within each episode, and the resistance burden index summarized the cumulative antibiotic resistance burden across antibiotics.

The action space consisted of three discrete control actions: decrease copper concentration, maintain the current level, or increase copper concentration. These actions corresponded to fixed stepwise adjustments in copper exposure applied at each cycle. Episodes were initialized at a sublethal copper concentration of 10 mg/L, consistent with exposure estimates reported by Li and colleagues [21], with copper levels constrained to remain within 0–100 mg/L throughout simulation.

Other biological dynamics, including changes in MIC values and growth inhibition, were modeled using stochastic update rules intended to represent resistance progression and partial reversibility under reduced antimicrobial pressure rather than explicit mechanistic mutation processes. To promote interpretability and reduce state-space complexity, continuous state variables were discretized into coarse bins prior to Q-learning. This design choice emphasized conceptual biological relevance and clarity over fine-grained mechanistic resolution, consistent with the exploratory objectives of the study.

2.3. Resistance Evolution

Consistent with prior work framing antimicrobial resistance as a feedback-control problem under noisy and partially observed dynamics [22], changes in antibiotic resistance were modeled using stochastic update rules intended to capture conceptual resistance progression under antimicrobial pressure rather than explicit molecular or genetic mechanisms. At each cycle t, resistance amplification occurred with probability:

$$P_{amp}(t) = \frac{C_t}{100}$$

where C_t demotes copper concentration (mg/L). When amplification occurred, MIC values for chloramphenicol and polymyxin B were increased by bounded stochastic increments drawn from predefined uniform ranges. When copper concentration fell below a low-exposure threshold, small stochastic decrements were applied to represent partial phenotypic reversibility. These update rules are heuristic and intentionally simplified, providing a transparent control environment rather than modeling mutation rates or population genetic processes.

2.4. Resistance Burden Index

The resistance burden index (RBI) was defined deterministically as:

$$RBI = 1 - \frac{MIC_{chloro} + MIC_{polyB}}{MIC_{Max}}$$

where MIC_{Max} represents the maximum possible summed MIC value across antibiotics (512 + 512 = 1024). This normalization constrains the index to the interval [0, 1], where higher values correspond to lower aggregate resistance burden and lower values reflect increased resistance. The index is therefore a scaled transformation of MIC values and does not represent empirical growth kinetics or independent biological signal.

2.5. Reward Function

Agent learning was guided by a scalar reward function designed to balance resistance suppression and copper minimization. Rewards increased with lower aggregate resistance burden, as reflected by the resistance burden index, and were penalized by elevated MIC values and increased copper exposure. Weighting terms were selected to keep individual reward components on comparable scales and to discourage trivial strategies that rely on persistently high copper concentrations. Reward parameters were chosen to reflect qualitative trade-offs rather than to maximize performance through extensive tuning, consistent with the exploratory aims of the study.

2.6. Experiment Protocol and Data Generation

Experiments were conducted in a fixed-length episodic simulation setting. Each episode began with a reset of the environment to baseline conditions and proceeded for a maximum of 40 cycles. At each cycle, an agent selected one of three discrete copper-control actions (decrease, maintain, increase), after which the environment updated copper concentration, minimum inhibitory concentration (MIC) values for chloramphenicol and polymyxin B, the growth inhibition proxy, and reward. Per-cycle outcomes were recorded for each episode, including copper concentration, MIC trajectories for both antibiotics, resistance burden index values, and reward.

The random-policy and rule-based agents were evaluated over 50 independent episodes each. The Q-learning agent was run for 100 episodes with online temporal-difference updates applied at each cycle. Because the Q-learning agent updates its policy across episodes, it was run for 100 episodes to permit cumulative adaptation. In contrast, the random and rule-based agents do not update their policies and therefore were evaluated over 50 independent episodes each. Performance metrics for all agents were computed by aggregating per-cycle outcomes across their respective simulation runs. As such, differences in cumulative exposure reflect adaptive learning behavior rather than unequal evaluation conditions. The Q-learning agent reflects adaptive control through

online learning, whereas the random and rule-based agents are non-learning baselines with fully specified policies.

2.7. Random Policy Agent

The random policy agent served as a negative-control baseline and selected copper-control actions uniformly at random at each cycle. This agent did not incorporate any information from the environment state and did not adapt its behavior based on observed resistance dynamics or prior outcomes. As such, the random policy represents an uncontrolled exposure scenario in which copper adjustments are applied without feedback or strategic intent.

Performance of the random policy reflects baseline resistance trajectories under non-optimized copper exposure and provides a reference point for evaluating the added value of structured and adaptive decision-making strategies.

2.8. Rule-Based Agent

The rule-based agent followed predefined decision rules to adjust copper exposure based on observed antibiotic resistance, as measured by MIC values. Copper levels were reduced when resistance increased, maintained when MIC values remained stable, and increased when resistance decreased. This agent reflects a protocol-driven decision strategy that could be applied in settings with limited computational or diagnostic resources. Its performance was compared with that of the Q-learning agent to evaluate the potential benefits of adaptive strategies for managing resistance dynamics.

2.9. Q-Learning Agent

The Q-learning agent implemented a tabular, model-free reinforcement learning approach to learn an adaptive copper-control policy through repeated interaction with the simulation environment. Because the environment state variables are continuous, the agent used coarse discretization to map each state to a finite set of bins prior to learning. The resulting discrete state space contains 10,000 states (10^4), yielding a Q-table of 30,000 state–action values (10,000 $\times$ 3 actions), a computationally modest scale that does not introduce meaningful sparsity constraints within the episodic training framework. Specifically, copper concentration (0–100 mg/L), MIC values for chloramphenicol and polymyxin B (0–512), and cycle count (0–max cycles) were each discretized into 10 equal-width bins and encoded as a four-dimensional discrete state index [27]. The resistance burden index was retained for reward computation but was not included in the Q-table state representation.

Action selection followed an ε-greedy policy ($\varepsilon = 0.10$): with probability ε the agent sampled a random action, otherwise it selected the action with the highest estimated state–action value. Q-values were updated online at each cycle using the standard temporal-difference update rule with learning rate $\alpha = 0.10$ and discount factor $\gamma = 0.95$. The Q-table persisted across episodes, allowing cumulative learning over repeated simulation runs. To evaluate robustness of the learned policy to parameter specification, additional hyperparameter and reward sensitivity analyses are provided in the Supplementary Materials (Tables S1–S3).

2.10. Longitudinal Modeling of Resistance Trajectories

To quantify differences in antimicrobial resistance trajectories across agent strategies over repeated simulation cycles, longitudinal statistical models were applied to the per-cycle simulation outputs. In addition to summary performance metrics, including the area under the curve (AUC) of minimum inhibitory concentration (MIC) values, cumulative copper burden, and mean resistance burden index value, cumulative copper burden, and

mean growth inhibition, these models were used to directly evaluate temporal patterns in resistance evolution across agents.

Resistance outcomes for chloramphenicol and polymyxin B were analyzed using generalized linear mixed models (GLMMs) implemented in R (packages: lme4, lmerTest, readr, and emmeans) [28]. Mixed-effects modeling was selected to account for the repeated-measures structure [29,30] of the simulation data, in which MIC values were recorded across multiple cycles within each episode. This approach allows separation of systematic effects attributable to agent behavior from stochastic variability across simulation runs.

For each antibiotic, MIC values were modeled as a function of simulation cycle, agent type (random, rule-based, or Q-learning), and their interaction, with episode included as a random intercept to capture between-episode heterogeneity. The inclusion of cycle-by-agent interaction terms enabled direct comparison of resistance trajectories over time, rather than reliance on single time-point or aggregate measures alone.

Model results are reported as estimated effects with 95% confidence intervals (CIs), providing interpretable measures of uncertainty and effect magnitude. Pairwise contrasts between agent types were evaluated using marginal means with multiplicity-adjusted confidence intervals to assess meaningful differences in resistance dynamics across control strategies. Together, these analyses complement AUC-based summaries by providing a model-based assessment of how resistance trajectories diverge under different decision-making frameworks.

2.11. Reproducibility and Generative AI Statement

All code used for experiments, data analysis, and figure generation is provided in the Supplemental Materials S1. Scripts are organized with clear instructions for setup and execution. Raw data and generated figures are saved in CSV and image formats, respectively, allowing validation and extension of the framework. Generative AI tools were used for coding assistance, troubleshooting, and initial idea generation; all outputs were reviewed prior to inclusion in the final manuscript.

3. Results

3.1. Agent-Level Resistance Dynamics Based on AUC Metrics

Across both antibiotics, the Q-learning agent exhibited lower cumulative resistance exposure than the random and rule-based agents, as measured by the area under the curve (AUC) of minimum inhibitory concentration (MIC) values (Tables 1 and 2 and Figures 1–4). Bootstrap contrasts indicated lower AUC MIC values for the Q-learning agent relative to the random agent for both chloramphenicol and polymyxin B. Differences between the Q-learning and rule-based agents were smaller, consistent with intermediate performance under rule-based decision-making (Table 1).

These patterns were reflected in aggregate performance metrics across experimental runs. The Q-learning agent achieved lower mean AUC MIC values for both antibiotics and lower copper burden compared with the random and rule-based agents (Table 2). In contrast, the random agent exhibited higher cumulative MIC exposure and copper burden, with greater variability across runs, while the rule-based agent demonstrated intermediate values across metrics Mean resistance burden index values remained near their upper bound across agents, indicating that between-agent differences primarily reflected divergence in resistance trajectories rather than short-term shifts in resistance burden. This suggests that differences in AUC MIC primarily reflected divergence in resistance trajectories over time rather than short-term differences in growth suppression.

Table 1. Bootstrap AUC summary for agent comparisons.

Contrast	Metric	Median Difference	95% CI
Q-Learner vs. Random	AUC MIC (Chloro)	−114.818	−195.987–−43.559
Q-Learner vs. Random	AUC MIC (PolyB)	−102.130	−153.837–−44.596
Q-Learner vs. Rule-Based	AUC MIC (Chloro)	−41.116	−54.349–−32.529
Q-Learner vs. Rule-Based	AUC MIC (PolyB)	−50.967	−58.493–−43.016
Rule-Based vs. Random	AUC MIC (Chloro)	−58.424	−141.439–−1.508
Rule-Based vs. Random	AUC MIC (PolyB)	−44.177	−101.903–−2.793

Table 2. Summary metrics for agent performance (Mean and Standard Deviation) for Chloramphenicol (Chloro) and Polymyxin B (PolyB).

Agent	AUC MIC (Chloro) Mean (SD)	AUC MIC (PolyB) Mean (SD)	Copper Burden Mean (SD)	Mean Growth Inhibition Mean (SD)
Q-Learner	10.131 (8.917)	22.229 (12.035)	15.050 (9.087)	0.999 (0.000)
Random	183.340 (175.391)	149.517 (113.335)	564.200 (394.399)	0.992 (0.007)
Rule-Based	69.739 (52.945)	76.042 (32.698)	142.200 (91.390)	0.996 (0.002)

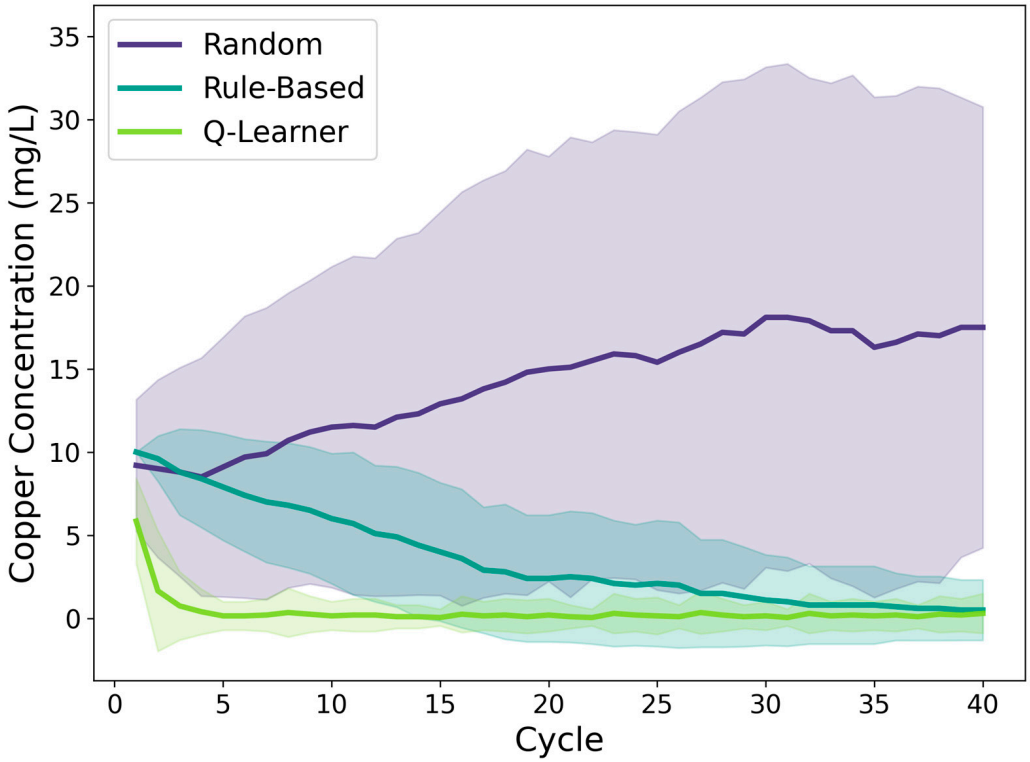


Figure 1. Copper concentration (mg/L) across experimental cycles for each agent. Lines represent the mean copper concentration across episodes for the random, rule-based, and Q-learning agents. Shaded regions denote ±1 standard deviation across episodes at each cycle.

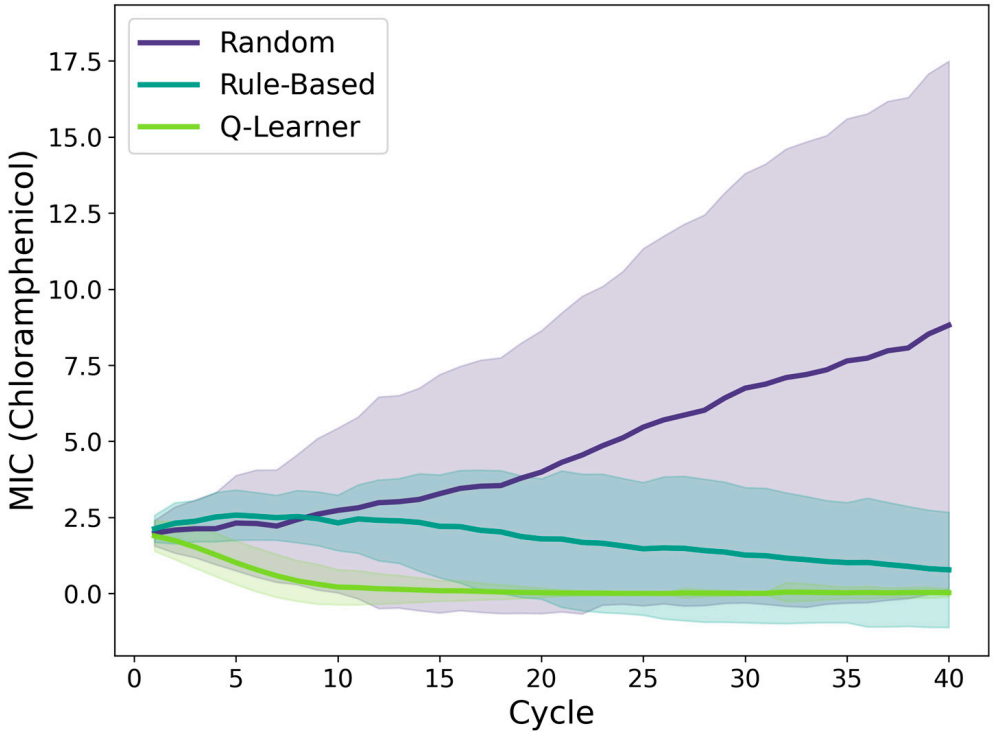


Figure 2. Chloramphenicol minimum inhibitory concentration (MIC) trajectories across experimental cycles. Lines represent mean MIC values across episodes for each agent. Shaded regions denote ±1 standard deviation across episodes, illustrating variability in resistance dynamics.

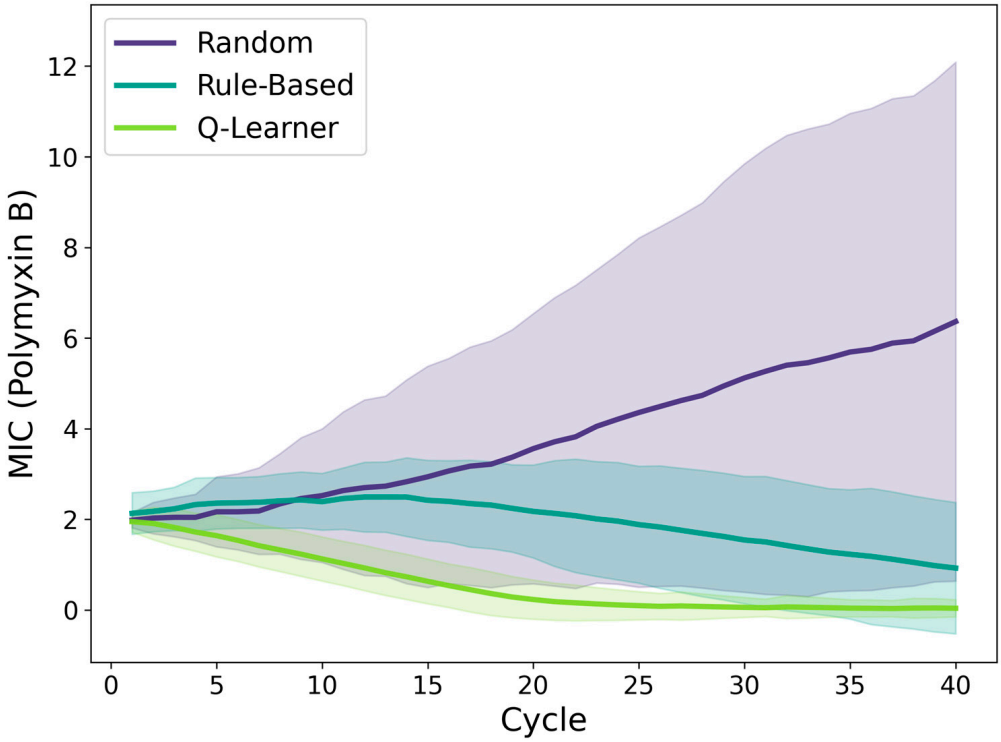


Figure 3. Polymyxin B minimum inhibitory concentration (MIC) across experimental cycles for the random, rule-based, and Q-learning agents. Lines represent mean MIC values across episodes, and shaded regions denote ±1 standard deviation across episodes.

https://doi.org/10.3390/applbiosci5020038

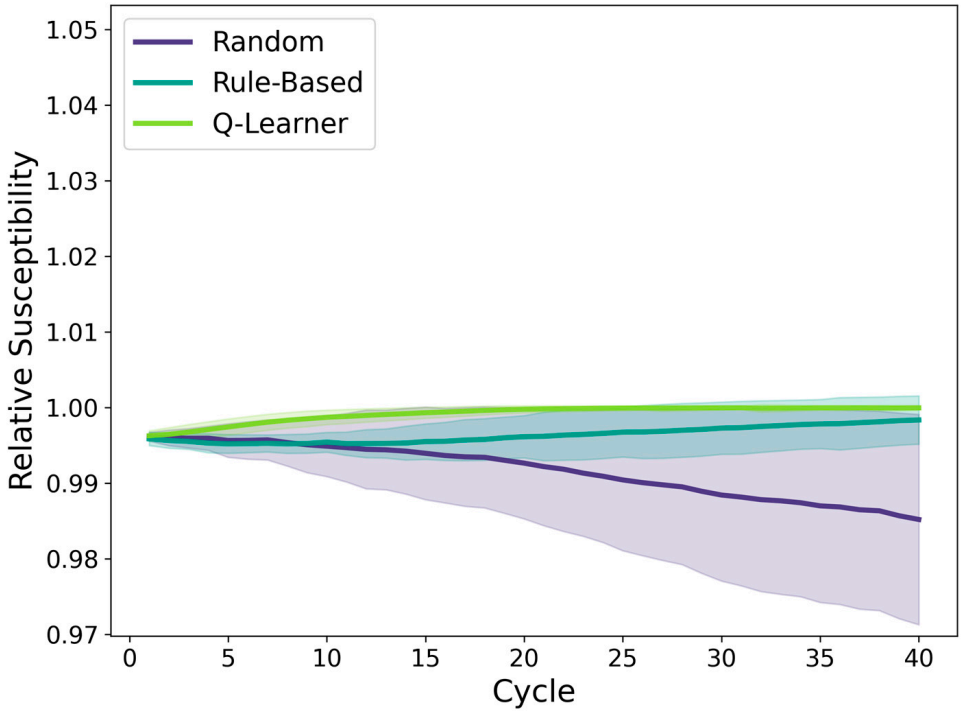


Figure 4. Resistance burden index across experimental cycles for the random, rule-based, and Q-learning agents. Lines represent mean values across episodes, with shaded regions indicating ±1 standard deviation at each cycle.

3.2. GLMM Analysis of Resistance Trajectories

Results from generalized linear mixed models (GLMMs) aligned with the AUC-based summaries, providing model-based estimates of differences in simulated resistance trajectories across agents (Table 3). Type III ANOVA tests indicated a strong main effect of agent and a significant cycle × agent interaction for both antibiotics ($p < 0.001$), with substantial marginal R^2 values across models. Relative to the Q-learning agent, the random and rule-based agents were associated with higher MIC values over cycles for both antibiotics, with effect sizes varying by agent and antibiotic. Interaction terms between cycle and agent indicated differential temporal patterns in MIC trajectories. The random agent showed increasing MIC values over cycles relative to the Q-learning agent, while the rule-based agent exhibited smaller interaction effects, consistent with intermediate resistance trajectories. Across models, estimated effects for polymyxin B were generally smaller in magnitude than those observed for chloramphenicol, although the direction of effects was consistent across antibiotics.

Table 3. Changes in antibiotic minimum inhibitory concentration (MIC) across simulation cycles by agent type, with 95% confidence intervals.

Effect	Chloramphenicol Estimate (95% CI)	Polymyxin B Estimate (95% CI)
Cycle	−0.031 (−0.038–−0.024)	−0.051 (−0.055–−0.046)
Random	0.062 (−0.229–0.352)	−0.222 (−0.412–−0.033)
Rule-Based	1.879 (1.589–2.169)	1.113 (0.924–1.302)
Cycle × Random	0.214 (0.202–0.226)	0.170 (0.162–0.178)
Cycle × Rule-Based	−0.018 (−0.030–−0.006)	0.012 (0.004–0.020)

4. Discussion

This study explored the application of reinforcement learning (RL) to model explore strategies for modulating antimicrobial resistance dynamics (AMR) under copper-driven antimicrobial pressure. Across simulation experiments, the Q-learning agent consistently achieved lower cumulative resistance exposure and copper burden compared to both rule-based and random policies, demonstrating improved performance of adaptive decision-making within this simulated resistance management setting. These findings align with prior work framing AMR as a feedback-control problem, where adaptive policies have shown improved performance relative to static or heuristic strategies within controlled modeling environments [22,24].

While the observed performance differences across agents were robust at the simulation level, this framework was intentionally designed as an exploratory model rather than a mechanistic representation of microbial evolution. As such, further refinement and empirical validation would be required to capture the full biological complexity of resistance dynamics observed in natural and clinical settings, including spatial structure, population heterogeneity, and ecological interactions [17,31,32]. The present results should therefore be interpreted as illustrating relative agent behavior under controlled assumptions rather than predictive estimates of resistance outcomes.

This study establishes a baseline framework for applying reinforcement learning approaches to non-antibiotic antimicrobial stressors, using copper as a representative compound with well-documented antimicrobial properties [15–18]. In contrast to prior RL-based AMR studies that focus on synthetic antibiotics and empirically measured fitness landscapes [22], the present work emphasizes interpretability, limited state information, and simplified control actions. This design choice reflects scenarios in which detailed molecular diagnostics or high-resolution resistance measurements may be unavailable, such as resource-limited clinical or agricultural environments [33,34].

Several limitations of the current framework warrant explicit consideration. First, resistance evolution was modeled using stochastic update rules rather than explicit genetic or molecular mechanisms. In addition, the modeled state variables were intentionally limited to MIC values and copper concentration and do not incorporate ecological factors such as population density, growth kinetics, or spatial structure. While this abstraction is consistent with prior systems-level approaches to AMR modeling [22,24], it does not capture processes such as horizontal gene transfer, compensatory mutations, or multi-species interactions, all of which play critical roles in real-world resistance dynamics [35,36]. Additionally, resistance burden was represented using an index derived from MIC values rather than empirical growth measurements or fitness assays, limiting direct biological interpretation of growth or viability outcomes. Third, copper was treated as a controllable environmental stressor rather than a therapeutic intervention, and the framework does not address toxicity, pharmacokinetics, or host-mediated effects that would be relevant in clinical translation [17,37,38].

Despite these limitations, the framework provides a useful platform for examining how differing levels of adaptability influence resistance trajectories under antimicrobial pressure. where adaptive policies have shown improved performance relative to static or heuristic strategies within controlled modeling environments. Within the context of this modeling framework, reinforcement learning provides a flexible methodological lens for studying AMR as a dynamic process shaped by feedback between microbial populations, antimicrobial exposure, and environmental constraints [22,24,33,34,39]. Furthermore, the present implementation focuses on a single representative metal stressor to preserve interpretability and controlled comparison of decision strategies; extension to additional metals

or antimicrobial agents would require expanded parameterization and is acknowledged as future work.

Future work could extend this framework by incorporating additional sources of biological complexity, including multi-species populations, additional metals, spatial structure, and horizontal gene transfer mechanisms [35,36]. Further studies could also evaluate alternative non-antibiotic antimicrobial compounds or combinations thereof, as well as more advanced learning algorithms capable of handling continuous state spaces and delayed rewards. Rather than providing a definitive solution to antimicrobial resistance, this study contributes an interpretable and extensible modeling framework that bridges microbiology and adaptive decision-making, laying groundwork for future investigations into AI-assisted AMR control strategies.

5. Conclusions

This study presented an exploratory reinforcement learning (RL) framework for modeling antimicrobial resistance (AMR) dynamics under copper-driven antimicrobial pressure. By comparing a random policy, a rule-based heuristic, and a Q-learning agent within a controlled simulation environment, this work demonstrates how differing levels of adaptability influence resistance trajectories over time. Across experiments, the Q-learning agent consistently achieved lower cumulative resistance exposure and copper burden than non-adaptive or fixed-decision strategies, illustrating the potential advantages of adaptive control within the modeled copper-driven resistance dynamics.

Importantly, the framework was designed to emphasize interpretability and conceptual clarity rather than biological or computational complexity. Resistance evolution and resistance burden were modeled using simplified, stochastic dynamics grounded in commonly reported microbiological indicators, such as minimum inhibitory concentration (MIC), without explicitly modeling molecular mechanisms. This design choice reflects scenarios in which detailed genetic data or high-resolution diagnostics may be unavailable, while still allowing meaningful comparison of control strategies under antimicrobial pressure.

Rather than proposing copper as a therapeutic intervention, this study uses copper as a representative non-antibiotic antimicrobial stressor to explore broader methodological questions at the intersection of microbiology and adaptive decision-making. The results indicate that, within the present simulation design, reinforcement learning approaches outperformed static or heuristic policies across repeated exposure cycles. These findings support the use of RL as a flexible modeling tool for studying AMR as a dynamic, feedback-driven process. The value of this framework lies not in biological realism, but in providing a transparent testbed for comparing decision strategies under shared and explicitly constrained assumptions.

The framework introduced here is intentionally modular and extensible. Future work may incorporate additional biological complexity, such as multi-species interactions, spatial structure, or alternative antimicrobial compounds, as well as more advanced learning algorithms capable of operating in continuous or partially observed environments. In parallel, empirical validation and tighter integration with experimental data would be necessary steps toward translational relevance.

Taken together, this work introduces a transparent and extensible simulation framework integrating microbiological indicators with reinforcement learning, establishing a foundation for future methodological research on antimicrobial resistance dynamics under constrained or uncertain conditions. These findings are specific to the modeled copper-based resistance environment and should not be generalized beyond the scope of the present simulation framework.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/applbiosci5020038/s1. All relevant code, data, and figures are available in Supplemental Materials S1. Additional hyperparameter and reward sensitivity analyses are provided in Supplemental Materials S2 (Table S1. Q-Learning Hyperparameter Sensitivity Results; Table S2. Ranked Q-Learning Hyperparameter Configurations; Table S3. Reward Function Sensitivity Results).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All simulation code, generated data, and analysis scripts supporting this study are publicly available in the project repository at: https://github.com/h-hedman/healthcare-data-science/tree/main/organic-amr-rl (accessed on 15 December 2025). Additional documentation describing the repository structure and reproducibility workflow is provided in Supplemental Materials S2.

Acknowledgments: The author thanks the reviewers for their constructive feedback, which helped improve the clarity of the manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Oldenkamp, R.; Schultsz, C.; Mancini, E.; Cappuccio, A. Filling the gaps in the global prevalence map of clinical antimicrobial resistance. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2013515118. [CrossRef]
2. Vikesland, P.; Garner, E.; Gupta, S.; Kang, S.; Maile-Moskowitz, A.; Zhu, N. Differential Drivers of Antimicrobial Resistance across the World. *Acc. Chem. Res.* **2019**, *52*, 916–924. [CrossRef] [PubMed]
3. Bertagnolio, S.; Dobreva, Z.; Centner, C.M.; Olaru, I.D.; Donà, D.; Burzo, S.; Huttner, B.D.; Chaillon, A.; Gebreselassie, N.; Wi, T.; et al. WHO global research priorities for antimicrobial resistance in human health. *Lancet Microbe* **2024**, *5*, 100902. [CrossRef] [PubMed]
4. Muteeb, G.; Rehman, M.T.; Shahwan, M.; Aatif, M. Origin of Antibiotics and Antibiotic Resistance, and Their Impacts on Drug Development: A Narrative Review. *Pharmaceuticals* **2023**, *16*, 1615. [CrossRef]
5. Chiș, A.A.; Rus, L.L.; Morgovan, C.; Arseniu, A.M.; Frum, A.; Vonica-Țincu, A.L.; Gligor, F.G.; Mureșan, M.L.; Dobrea, C.M. Microbial Resistance to Antibiotics and Effective Antibiotherapy. *Biomedicines* **2022**, *10*, 1121. [CrossRef]
6. Marston, H.D.; Dixon, D.M.; Knisely, J.M.; Palmore, T.N.; Fauci, A.S. Antimicrobial Resistance. *JAMA* **2016**, *316*, 1193–1204. [CrossRef]
7. Brüssow, H. The antibiotic resistance crisis and the development of new antibiotics. *Microb. Biotechnol.* **2024**, *7*, e14510. [CrossRef] [PubMed]
8. Iskandar, K.; Murugaiyan, J.; Halat, D.H.; Hage, S.E.; Chibabhai, V.; Adukkadukkam, S.; Roques, C.; Molinier, L.; Salameh, P.; Van Dongen, M. Antibiotic Discovery and Resistance: The Chase and the Race. *Antibiotics* **2022**, *11*, 182. [CrossRef]
9. Elshobary, M.E.; Badawy, N.K.; Ashraf, Y.; Zatioun, A.A.; Masriya, H.H.; Ammar, M.M.; Mohamed, N.A.; Mourad, S.; Assy, A.M. Combating Antibiotic Resistance: Mechanisms, Multidrug-Resistant Pathogens, and Novel Therapeutic Approaches: An Updated Review. *Pharmaceuticals* **2025**, *18*, 402. [CrossRef]
10. Berger, I.; Loewy, Z.G. Antimicrobial Resistance and Novel Alternative Approaches to Conventional Antibiotics. *Bacteria* **2024**, *3*, 171–182. [CrossRef]
11. Elfadadny, A.; Ragab, R.F.; AlHarbi, M.; Badshah, F.; Ibáñez-Arancibia, E.; Farag, A.; Hendawy, A.O.; De los Rios-Escalante, P.R.; Aboubakr, M.; Zakai, S.A.; et al. Antimicrobial resistance of Pseudomonas aeruginosa: Navigating clinical impacts, current resistance trends, and innovations in breaking therapies. *Front. Microbiol.* **2024**, *15*, 1374466. [CrossRef]
12. Bisht, N.; Dwivedi, N.; Kumar, P.; Venkatesh, M.; Yadav, A.K.; Mishra, D.; Salanki, P.; Verma, N.K.; Lakshminarayanan, R.; Ramakrishna, S.; et al. Recent advances in copper and copper-derived materials for antimicrobial resistance and infection control. *Curr. Opin. Biomed. Eng.* **2022**, *24*, 100408. [CrossRef] [PubMed]
13. Virieux-Petit, M.; Hammer-Dedet, F.; Aujoulat, F.; Jumas-Bilak, E.; Romano-Bertrand, S. From Copper Tolerance to Resistance in Pseudomonas aeruginosa towards Patho-Adaptation and Hospital Success. *Genes* **2022**, *13*, 301. [CrossRef] [PubMed]

14. Yan, L.; Gopal, A.; Kashif, S.; Hazelton, P.; Lan, M.; Zhang, W.; Chen, X. Metal organic frameworks for antibacterial applications. *Chem. Eng. J.* **2022**, *435*, 134975. [CrossRef]
15. Ivanova, I.A.; Daskalova, D.S.; Yordanova, L.P.; Pavlova, E.L. Copper and Copper Nanoparticles Applications and Their Role against Infections: A Minireview. *Processes* **2024**, *12*, 352. [CrossRef]
16. Wang, Y.; Wen, T.; Mao, F.; Yang, S.; Zhang, Q.; Fu, X.; Zhai, C.; Zhang, H. Engineering copper and copper-based materials for a post-antibiotic era. *Front. Bioeng. Biotechnol.* **2025**, *13*, 1644362. [CrossRef]
17. Frei, A.; Verderosa, A.D.; Elliott, A.G.; Zuegg, J.; Blaskovich, M.A.T. Metals to combat antimicrobial resistance. *Nat. Rev. Chem.* **2023**, *7*, 202–224. [CrossRef]
18. Yu, J.; Huang, X.; Ren, F.; Cao, H.; Yuan, M.; Ye, T.; Xu, F. Application of antimicrobial properties of copper. *Appl. Organomet. Chem.* **2024**, *38*, e7506. [CrossRef]
19. Liu, W.; Xu, Y.; Slaveykova, V.I. Oxidative stress induced by sub-lethal exposure to copper as a mediator in development of bacterial resistance to antibiotics. *Sci. Total Environ.* **2023**, *860*, 160516. [CrossRef]
20. Forouzandeh, A.; Lassen, S.B.; Brinck, J.E.; Zhou, Y.-Y.; Zhu, J.; Solà-Oriol, D.; Monteiro, A.; Hao, X.; Su, J.Q.; Stein, H.H.; et al. Limited impacts of high doses of dietary copper on the gut bacterial metal resistome explain negligible co-selection of antibiotic resistance. *Sci. Total Environ.* **2023**, *889*, 164183. [CrossRef]
21. Li, J.; Phulpoto, I.A.; Zhang, G.; Yu, Z. Acceleration of emergence of E. coli antibiotic resistance in a simulated sublethal concentration of copper and tetracycline co-contaminated environment. *AMB Express* **2021**, *11*, 14. [CrossRef]
22. Weaver, D.T.; King, E.S.; Maltas, J.; Scott, J.G. Reinforcement learning informs optimal treatment strategies to limit antibiotic resistance. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2303165121. [CrossRef]
23. Rawson, T.M.; Moore, L.S.P. Understanding how diagnostics influence antimicrobial decision-making is key to successful clinical trial design. *Clin. Microbiol. Infect.* **2023**, *29*, 666–669. [CrossRef]
24. Tomurcuk, A.E.; Bozcali, E.; Baytemur, F.D.; Sivri, C.; Soyak, E.G. A Reinforcement Learning-inspired Estimation of Antibiotic Resistance in Poultry Environments. In *Engineering and Technology Management in Challenging Times*; Çebi, F., Tekin, A.T., Çorum, A., Cömert, G.D., Bolatan, G.I.S., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 91–99. [CrossRef]
25. Kadeřábková, N.; Mahmood, A.J.S.; Mavridou, D.A.I. Antibiotic susceptibility testing using minimum inhibitory concentration (MIC) assays. *npj Antimicrob. Resist.* **2024**, *2*, 37. [CrossRef]
26. Smaill, F. Antibiotic Susceptibility and Resistance Testing: An Overview. *Can. J. Gastroenterol. Hepatol.* **2000**, *14*, 382415. [CrossRef]
27. Spanò, S.; Cardarilli, G.C.; Di Nunzio, L.; Fazzolari, R.; Giardino, D.; Matta, M.; Nannarelli, A.; Re, M. An Efficient Hardware Implementation of Reinforcement Learning: The Q-Learning Algorithm. *IEEE Access* **2019**, *7*, 186340–186351. [CrossRef]
28. Bolker, B.M.; Brooks, M.E.; Clark, C.J.; Geange, S.W.; Poulsen, J.R.; Stevens, M.H.H.; White, J.S.S. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **2009**, *24*, 127–135. [CrossRef]
29. Hedman, H.D.; Eisenberg, J.N.S.; Trueba, G.; Rivera, D.L.V.; Herrera, R.A.Z.; Barrazueta, J.V.; Gavilanes Rodriguez, G.I.; Krawczyk, E.; Berrocal, V.J.; Zhang, L. Impacts of small-scale chicken farming activity on antimicrobial-resistant Escherichia coli carriage in backyard chickens and children in rural Ecuador. *One Health* **2019**, *8*, 100112. [CrossRef] [PubMed]
30. Amato, H.K.; Loayza, F.; Salinas, L.; Paredes, D.; Garcia, D.; Sarzosa, S.; Saraiva-Garcia, C.; Johnson, T.J.; Pickering, A.J.; Riley, L.W.; et al. Risk factors for extended-spectrum beta-lactamase (ESBL)-producing E. coli carriage among children in a food animal-producing region of Ecuador: A repeated measures observational study. *PLoS Med.* **2023**, *20*, e1004299. [CrossRef]
31. Denk-Lobnig, M.K.; Wood, K.B. Spatial population dynamics of bacterial colonies with social antibiotic resistance. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2417065122. [CrossRef] [PubMed]
32. Abel, K.; Agnew, E.; Amos, J.; Armstrong, N.; Armstrong-James, D.; Ashfield, T.; Aston, S.; Baillie, J.K.; Baldwin, S.; Barlow, G.; et al. System-wide approaches to antimicrobial therapy and antimicrobial resistance in the UK: The AMR-X framework. *Lancet Microbe* **2024**, *5*, e500–e507. [CrossRef]
33. DiazGranados, C.A.; Cardo, D.M.; McGowan, J.E. Antimicrobial resistance: International control strategies, with a focus on limited-resource settings. *Int. J. Antimicrob. Agents* **2008**, *32*, 1–9. [CrossRef] [PubMed]
34. Hijazi, K.; Joshi, C.; Gould, I.M. Challenges and opportunities for antimicrobial stewardship in resource-rich and resource-limited countries. *Expert Rev. Anti-Infect. Ther.* **2019**, *17*, 621–634. [CrossRef]
35. Wit, G.D.; Svet, L.; Lories, B.; Steenackers, H.P. Microbial Interspecies Interactions and Their Impact on the Emergence and Spread of Antimicrobial Resistance. *Annu. Rev. Microbiol.* **2022**, *76*, 179–192. [CrossRef] [PubMed]
36. Woods, L.C.; Gorrell, R.J.; Taylor, F.; Connallon, T.; Kwok, T.; McDonald, M.J. Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 26868–26875. [CrossRef]
37. Guedes, B.N.; Krambeck, K.; Durazzo, A.; Lucarini, M.; Santini, A.; Oliveira, M.B.P.P.; Fathi, F.; Suoto, E.B. Natural antibiotics against antimicrobial resistance: Sources and bioinspired delivery systems. *Braz. J. Microbiol.* **2024**, *55*, 2753–2766. [CrossRef] [PubMed]

38. Alaoui Mdarhri, H.; Benmessaoud, R.; Yacoubi, H.; Seffar, L.; Guennouni Assimi, H.; Hamam, M.; Boussettine, R.; Filali-Ansari, N.; Azzahra Lahlou, F.; Diawara, I.; et al. Alternatives Therapeutic Approaches to Conventional Antibiotics: Advantages, Limitations and Potential Application in Medicine. *Antibiotics* **2022**, *11*, 1826. [CrossRef]
39. Irfan, M.; Almotiri, A.; AlZeyadi, Z.A. Antimicrobial Resistance and Its Drivers—A Review. *Antibiotics* **2022**, *11*, 1362. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.