



## Computational Modelling and AI-Based Simulation of Host–Pathogen Interactions in Infectious Diseases

Emmanuel Nkansah<sup>1</sup>, Micheal Abimbola Oladosu<sup>2\*</sup>, Moses Adondua Abah<sup>3</sup>, Abimbola Mary Oluwajembola<sup>2</sup>, Fwangmun Ezekiel Gushit<sup>4</sup>, Olaide Ayokunmi Oladosu<sup>5</sup>, Adesola Esther Adeneye<sup>6</sup> and Bukola Oluwaseyi Olufosoye<sup>7</sup>

<sup>1</sup>Department of Accounting, Economics and Finance, School of Business, La Sierra University, Riverside, CA, USA

<sup>2</sup>Department of Chemical Sciences, Faculty of Science, Anchor University, Ayobo, Ipaja, Lagos, Nigeria

<sup>3</sup>Department of Biochemistry, Faculty of Pure and Applied Sciences, Federal University of Wukari, Wukari, Taraba State, Nigeria

<sup>4</sup>Department of Public Health, Faculty of Health Science, Ahmadu Bello University, Zaria, Kaduna State, Nigeria

<sup>5</sup>Department of Computer Science, Faculty of Science and Technology, Babcock University, Ilishan, Nigeria

<sup>6</sup>Department of Biological Sciences, Faculty of Science, Anchor University, Ayobo, Ipaja, Lagos, Nigeria

<sup>7</sup>Department of Medical Microbiology, Faculty of Medical Laboratory Sciences, Ambrose Alli University, Ekpoma, Edo State, Nigeria

**\*Corresponding Author:** Micheal Abimbola Oladosu, Department of Chemical Sciences, Faculty of Science, Anchor University, Ayobo, Ipaja, Lagos, Nigeria.

**DOI:** 10.31080/ASMI.2026.09.1593

### Abstract

The emergence of artificial intelligence (AI) and advanced computational methods has revolutionised our understanding of host–pathogen interactions in infectious diseases. This review comprehensively examines recent developments in computational modelling, machine learning algorithms, and AI-based simulation techniques applied to predicting and analysing molecular interactions between pathogens and their hosts. We discuss the integration of multi-omics data, protein–protein interaction prediction models, molecular dynamics simulations, and systems biology approaches that collectively enhance our capacity to identify therapeutic targets and understand infection mechanisms. Despite remarkable progress, challenges remain in data quality, model interpretability, and computational resource requirements. The synergistic application of AI with traditional experimental methods offers unprecedented opportunities for accelerating drug discovery, vaccine development, and precision medicine approaches against infectious diseases. This review highlights key methodologies, recent breakthroughs, and future directions in computational host–pathogen interaction research.

**Keywords:** Artificial Intelligence; Machine Learning; Host-Pathogen Interactions; Protein-Protein Interactions; Computational Modelling; Drug Discovery; Systems Biology; Molecular Dynamics

## Introduction

Infectious diseases continue to pose substantial threats to global public health, causing millions of deaths annually and imposing enormous economic burdens worldwide. The intricate molecular interactions between pathogens and their hosts determine infection outcomes, disease progression, and therapeutic responses. Traditional experimental approaches to studying these interactions, while essential, are often time-consuming, expensive, and limited in scope. The rapid advancement of computational technologies, particularly artificial intelligence (AI) and machine learning (ML), has opened new frontiers in infectious disease research, enabling researchers to predict, model, and understand host-pathogen interactions at unprecedented scales and speeds [1,2].

The COVID-19 pandemic dramatically highlighted the critical need for rapid computational tools that can accelerate our understanding of novel pathogens and facilitate therapeutic development. AI-powered approaches have demonstrated remarkable potential in various aspects of infectious disease research, from predicting viral protein structures using AlphaFold to identifying potential drug candidates through virtual screening platforms [3,4]. These computational methods leverage vast amounts of genomic, proteomic, and structural data to generate predictions that guide experimental validation and therapeutic interventions.

This review examines the current state of computational modelling and AI-based simulation in host-pathogen interaction research. We explore key methodologies including machine learning-based protein-protein interaction (PPI) prediction, molecular dynamics simulations, systems biology approaches, and network-based analysis. Furthermore, we discuss the integration of multi-omics data, challenges in model development, and emerging opportunities for translating computational predictions into clinical applications. Our focus emphasises recent developments from 2020-2025, a period marked by significant methodological

advances and the urgent need for rapid response to emerging infectious diseases.

## Computational approaches in host-pathogen research

### Machine learning for protein-protein interaction prediction

Protein-protein interactions form the molecular basis of host-pathogen relationships, mediating processes such as pathogen entry, immune evasion, and hijacking of host cellular machinery. Machine learning algorithms have emerged as powerful tools for predicting these interactions from sequence, structural, and functional data. Recent developments in ensemble learning methods, combining multiple ML classifiers such as support vector machines (SVM), random forests, and logistic regression, have achieved remarkable accuracy in PPI prediction [5,6].

The HPiP (Host-Pathogen Interaction Prediction) package exemplifies this approach, utilising amino acid sequence descriptors and ensemble ML classifiers to predict unmapped interactions between pathogen and host proteins. When tested against SARS-CoV-2 and Mycobacterium tuberculosis proteins, HPiP demonstrated strong predictive performance, with experimental validation confirming numerous predictions<sup>6</sup>. Such tools accelerate the discovery of novel interaction partners and potential therapeutic targets without requiring extensive experimental screening.

Network centrality features extracted from host-pathogen interaction networks have proven particularly valuable for drug target classification. Analysis of network parameters, including eigenvector centrality, betweenness, and clustering coefficients, enables prioritisation of proteins critical to infection pathways. Random forest classifiers trained on these features achieved accuracies exceeding 99% in distinguishing drug targets from non-targets in cardiovascular disease-associated pathogens [7]. Table 1 presents the machine learning Algorithms for host-pathogen Interaction Prediction.

Algorithm Type	Key Features	Applications	Accuracy
Support Vector Machine (SVM)	Kernel-based classification, high-dimensional feature space	PPI prediction, drug target identification	85-95%
Random Forest	Ensemble decision trees, feature importance ranking	Network-based target prediction, classification	90-99%
Deep Neural Networks	Multi-layer architecture, automatic feature learning	Sequence analysis, structure prediction	88-96%
AlphaFold/RoseTTAFold	Transformer architecture, attention mechanisms, MSA processing	Protein complex structure prediction, interface modelling	92-98%
Ensemble Methods	Multiple classifier integration, weighted voting	Comprehensive HPI prediction, robust classification	93-99%

**Table 1:** Machine Learning Algorithms for Host-Pathogen Interaction Prediction [5-7,9,16,21].

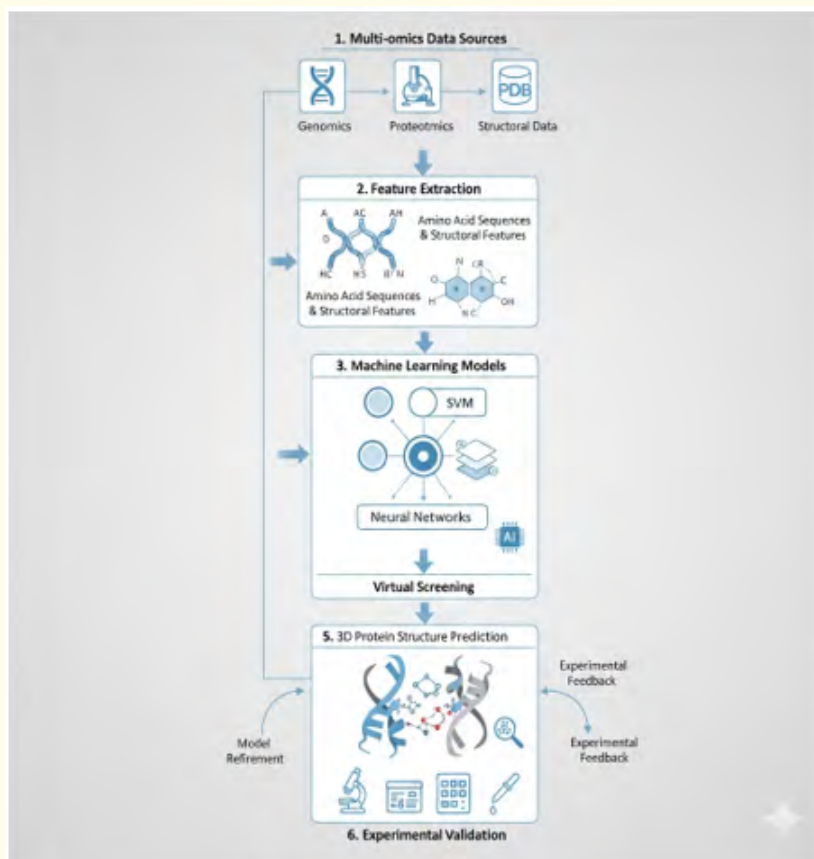
**Deep learning and structural prediction**

The advent of AI-powered structural prediction tools, particularly AlphaFold and RoseTTAFold, has revolutionised our ability to model protein complexes and predict interaction interfaces. These deep learning systems leverage residue-residue coevolution patterns and vast protein structure databases to generate highly accurate three-dimensional models. AlphaFold-Multimer has proven especially valuable for predicting protein complex structures, enabling systematic identification of host-pathogen interactions at proteome-wide scales [8,9].

A recent study employed RoseTTAFold2-Lite to screen 78 million protein pairs across 19 human bacterial pathogens, identifying 1,923 high-confidence complexes involving essential genes and 256

involving virulence factors. Many of these predictions represented previously unknown interactions, with experimental validation confirming approximately half of the tested predictions [9]. This demonstrates the power of deep learning approaches to uncover novel biological insights that would be practically impossible to discover through conventional experimental methods alone.

Integration of structural prediction with virtual drug screening platforms enables comprehensive pipelines for PPI-based drug discovery. One innovative approach combined AlphaFold-Multimer predictions with Virtual Flow, an ultra-large virtual screening platform, to identify inhibitors of the SARS-CoV-2 NSP10-NSP16 methyltransferase complex. This pipeline successfully identified a novel compound that inhibited viral replication, demonstrating the translational potential of AI-guided drug discovery [4].



**Figure 1:** Workflow of AI-Powered Host-Pathogen Interaction Prediction Pipeline [3,4,6]. The diagram illustrates the integration of multiple computational approaches from data acquisition through experimental validation. Key components include: (A) Multi-omics data collection from host and pathogen organisms, (B) Feature extraction and preprocessing using bioinformatics tools, (C) Machine learning model training with ensemble classifiers, (D) Protein structure prediction using AlphaFold-Multimer, (E) Molecular docking and virtual screening, and (F) Experimental validation cycle with feedback loop for model refinement. Colour coding indicates data flow between different analysis stages.

### Systems biology and network-based approaches

Systems biology approaches provide holistic frameworks for understanding infectious disease dynamics by integrating multi-omics data into comprehensive interaction networks. These methods recognise that biological systems exhibit emergent properties that cannot be predicted by studying individual components in isolation. Instead, network-based analysis reveals how pathogens exploit host signalling pathways, metabolic networks, and immune responses through coordinated molecular interactions [10,11].

Network modelling enables the identification of critical nodes and pathways that represent potential therapeutic targets. Analysis of host-pathogen interaction networks can reveal bottleneck proteins whose disruption would significantly impair pathogen

survival or replication. Moreover, integration of transcriptomic, proteomic, and metabolomic data allows construction of dynamic models that capture temporal changes during infection progression [12].

Recent advances in integrating AI with mechanistic epidemiological models have enhanced disease forecasting and intervention planning. A comprehensive scoping review identified 245 studies demonstrating how AI methods improve model parameterisation, calibration, and prediction accuracy. These integrated approaches bridge the gap between purely data-driven ML models and traditional mechanistic understanding, providing more robust and interpretable predictions [2]. Table 2 shows the systems biology approaches in host-pathogen research.

Approach	Data Types	Key Insights	Limitations
Multi-omics Integration	Genomics, transcriptomics, proteomics, metabolomics	Holistic view of infection dynamics	Data complexity
Network Biology	PPI networks, metabolic networks, and regulatory networks	Identification of key hubs and pathways	Network incompleteness
Agent-Based Models	Individual-level interactions, spatial dynamics	Emergent population-level behaviours	Computational cost
Mechanistic Models	Kinetic parameters, molecular interactions	Detailed molecular mechanisms	Parameter estimation
Hybrid AI-Mechanistic	Combined machine learning and mechanistic insights	Enhanced prediction and interpretability	Integration complexity

**Table 2:** Systems Biology Approaches in Host-Pathogen Research [2,10,11,22,27].

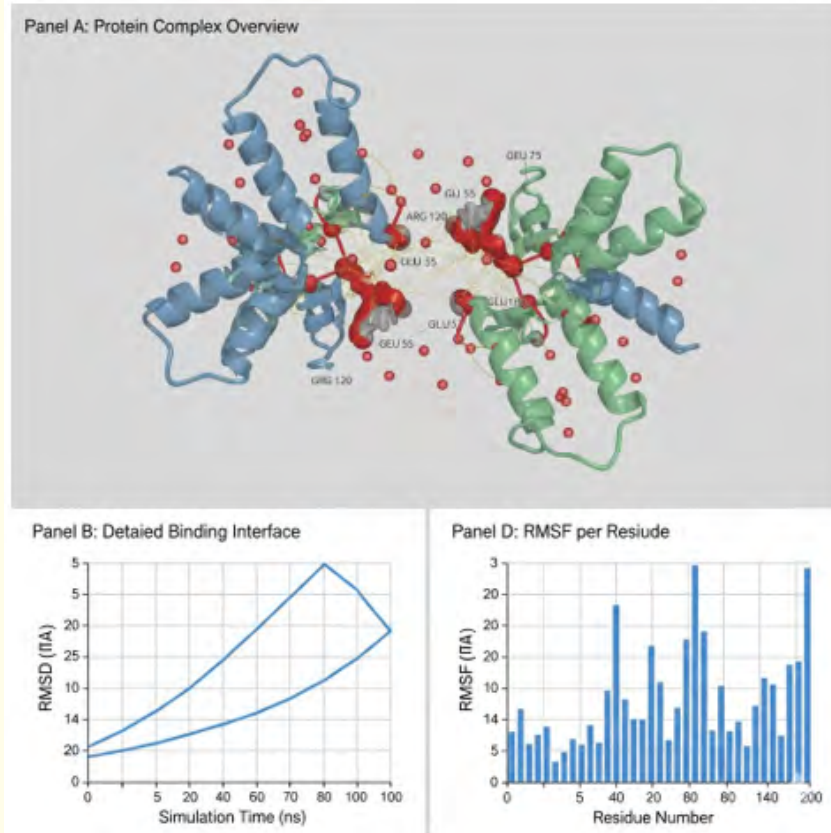
### Molecular dynamics simulations

Molecular dynamics (MD) simulations provide atomistic insights into the dynamic behaviour of host–pathogen protein complexes, revealing conformational changes, binding mechanisms, and interaction stability that static structures cannot capture. MD simulations employ physics-based force fields to model molecular motion over time, enabling researchers to observe how proteins interact, fold, and respond to environmental conditions [13,14].

Recent computational advances have enabled MD simulations of complete viral particles and large macromolecular assemblies. Coarse-grained approaches reduce computational complexity while maintaining biological relevance, allowing simulation of

systems containing millions of atoms over microsecond timescales. These simulations have proven instrumental in understanding viral entry mechanisms, identifying druggable pockets in pathogen proteins, and predicting how mutations affect protein stability and function [14].

Integration of MD simulations with machine learning creates powerful hybrid approaches. ML models trained on MD trajectories can predict binding affinities, identify stable interaction modes, and guide rational drug design. Furthermore, enhanced sampling techniques combined with AI-based analysis enable efficient exploration of conformational space, accelerating discovery of functionally important protein states [15].



**Figure 2:** Molecular Dynamics Simulation of Host-Pathogen Protein Complex [13-15]. Representative snapshot from a 100-nano-second MD simulation showing the SARS-CoV-2 spike protein receptor-binding domain (RBD) in complex with human ACE2 receptor. (A) Overall structure with RBD shown in blue and ACE2 in green, with key interface residues highlighted in red. (B) Close-up view of the binding interface showing critical hydrogen bonds (dashed yellow lines) and hydrophobic interactions. (C) Root-mean-square deviation (RMSD) plot over simulation time showing protein complex stability. (D) Per-residue root-mean-square fluctuation (RMSF) analysis identifying flexible and rigid regions. Interface water molecules are shown as red spheres. Scale bar represents 5 Å.

### Applications in drug discovery and vaccine development

The translation of computational predictions into therapeutic applications represents a critical frontier in infectious disease research. AI-guided drug discovery pipelines significantly accelerate the identification of promising compounds by computationally screening millions of molecules against predicted pathogen targets. Virtual screening platforms can evaluate binding affinity, selectivity, and drug-like properties *in silico*, dramatically reducing the experimental search space [16,17].

Machine learning models trained on experimental activity data can predict compound efficacy against specific pathogens, enabling

prioritisation of candidates for synthesis and testing. These approaches have proven particularly valuable for drug repurposing, where existing approved drugs are evaluated for activity against new pathogen targets. Computational screening identified several FDA-approved drugs with potent activity against SARS-CoV-2, leading to rapid clinical evaluation during the pandemic [18].

Vaccine development benefits from computational prediction of immunogenic epitopes and optimisation of vaccine candidates. AI models can predict T-cell and B-cell epitopes from pathogen protein sequences, identify conserved regions less prone to immune escape, and design multi-epitope vaccines. Structure-based design



### Challenges and limitations

Despite remarkable progress, computational approaches to host–pathogen interaction research face several significant challenges. Data quality and availability remain critical bottlenecks. Many pathogens lack comprehensive experimental data on protein structures, interactions, and functional annotations. Training robust ML models requires large, high-quality datasets, which are often unavailable for emerging or understudied pathogens [21,22].

Model interpretability presents another major challenge. Deep learning models often function as “black boxes,” making it difficult to understand why specific predictions are made or to identify when models might fail. This lack of transparency can hinder adoption in clinical settings where explainability is crucial. Furthermore, computational predictions must be experimentally validated, as false positives can waste resources and mislead research directions [23].

Computational resource requirements can be prohibitive, particularly for molecular dynamics simulations and large-scale virtual screening campaigns. While cloud computing and specialised hardware accelerate calculations, access remains unequal globally. Additionally, the rapid evolution of pathogens, particularly RNA viruses, poses challenges for models trained on historical data. Continuous model updating and adaptation to emerging variants demand sustained computational and human resources [24].

Ethical considerations surrounding AI in infectious disease research include data privacy, algorithmic bias, and equitable access to computational tools. Genomic and health data require robust protection, while ensuring that AI-driven discoveries benefit all populations, particularly in resource-limited settings where infectious disease burden is highest [1]. Table 3 highlights the challenges and future opportunities in AI-based host–pathogen research.

Category	Current Challenges	Future Opportunities
Data Quality	Incomplete datasets, biased sampling, annotation errors	Standardized data collection protocols, automated quality control
Model Interpretability	Black-box models, limited mechanistic understanding	Explainable AI methods, attention visualization, feature importance
Computational Resources	High computational costs, limited accessibility in low-resource settings	Cloud computing platforms, model optimization, hardware acceleration
Validation	Limited experimental validation, false positive predictions	Active learning, automated experimental workflows, closed-loop systems
Pathogen Evolution	Rapid mutation rates, emergence of variants, model obsolescence	Continuous learning models, evolutionary trajectory prediction, adaptive algorithms
Integration	Disconnect between molecular and clinical scales	Multi-scale modelling, clinical trial integration, precision medicine

**Table 3:** Challenges and Future Opportunities in AI-Based Host-Pathogen Research [1,2,12,17,21].

### Future directions and opportunities

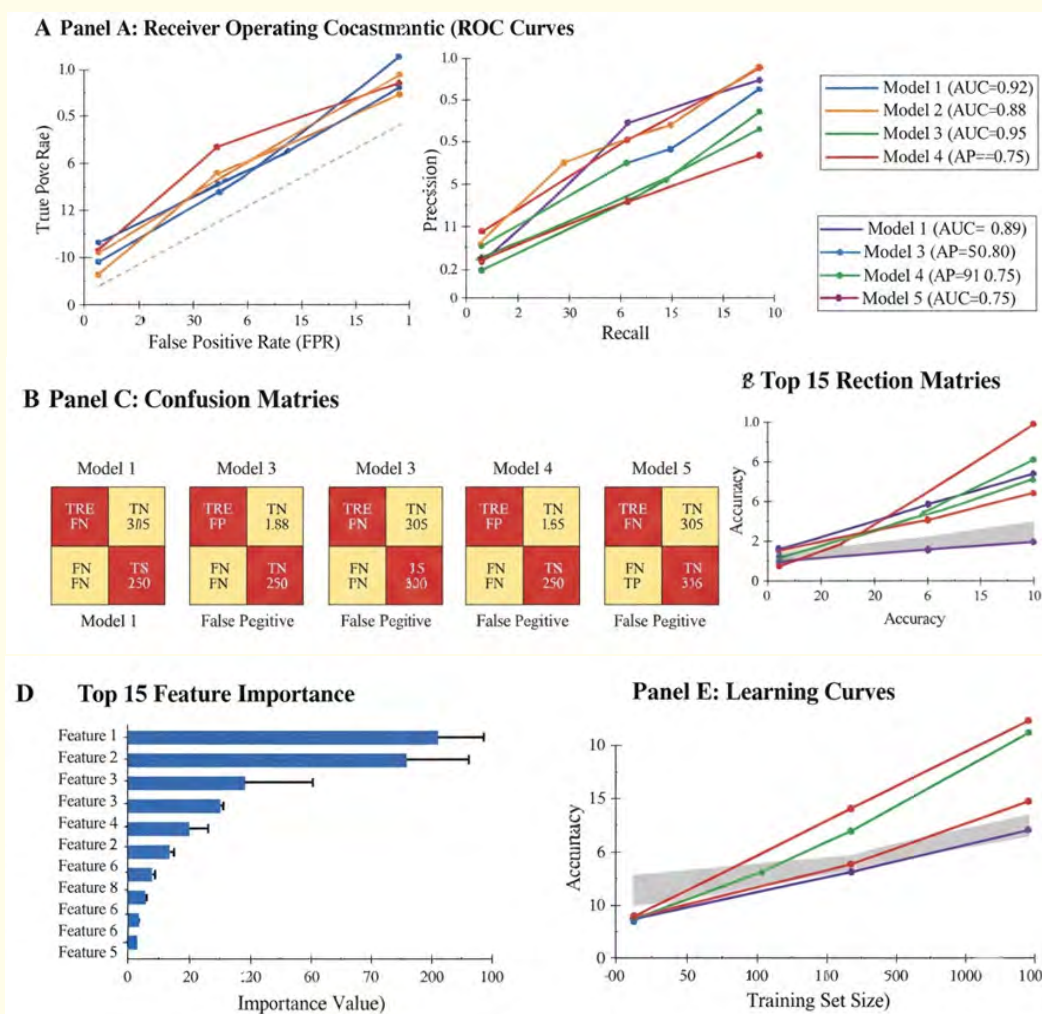
The future of computational host–pathogen interaction research lies in developing more integrated, interpretable, and accessible AI systems. Foundation models trained on diverse biological data show promise for transfer learning applications, enabling accurate predictions even with limited pathogen-specific data. These models can leverage knowledge from well-studied organisms to make informed predictions about novel pathogens [25].

Active learning approaches that iteratively combine computational predictions with experimental validation offer efficient strategies for exploring vast interaction spaces. Automation of experimental workflows through laboratory robotics can accelerate validation of computational predictions, creating closed-loop systems where AI guides experiments and experimental results continuously improve models [26].

Multi-scale modelling that connects molecular interactions to tissue-level and organism-level disease outcomes represents an exciting frontier. Integrating cellular models with epidemiological simulations could enable the prediction of how molecular interventions translate to population-level disease control. Agent-based models incorporating AI-predicted interaction networks may provide more realistic simulations of infection dynamics and intervention outcomes [27,28].

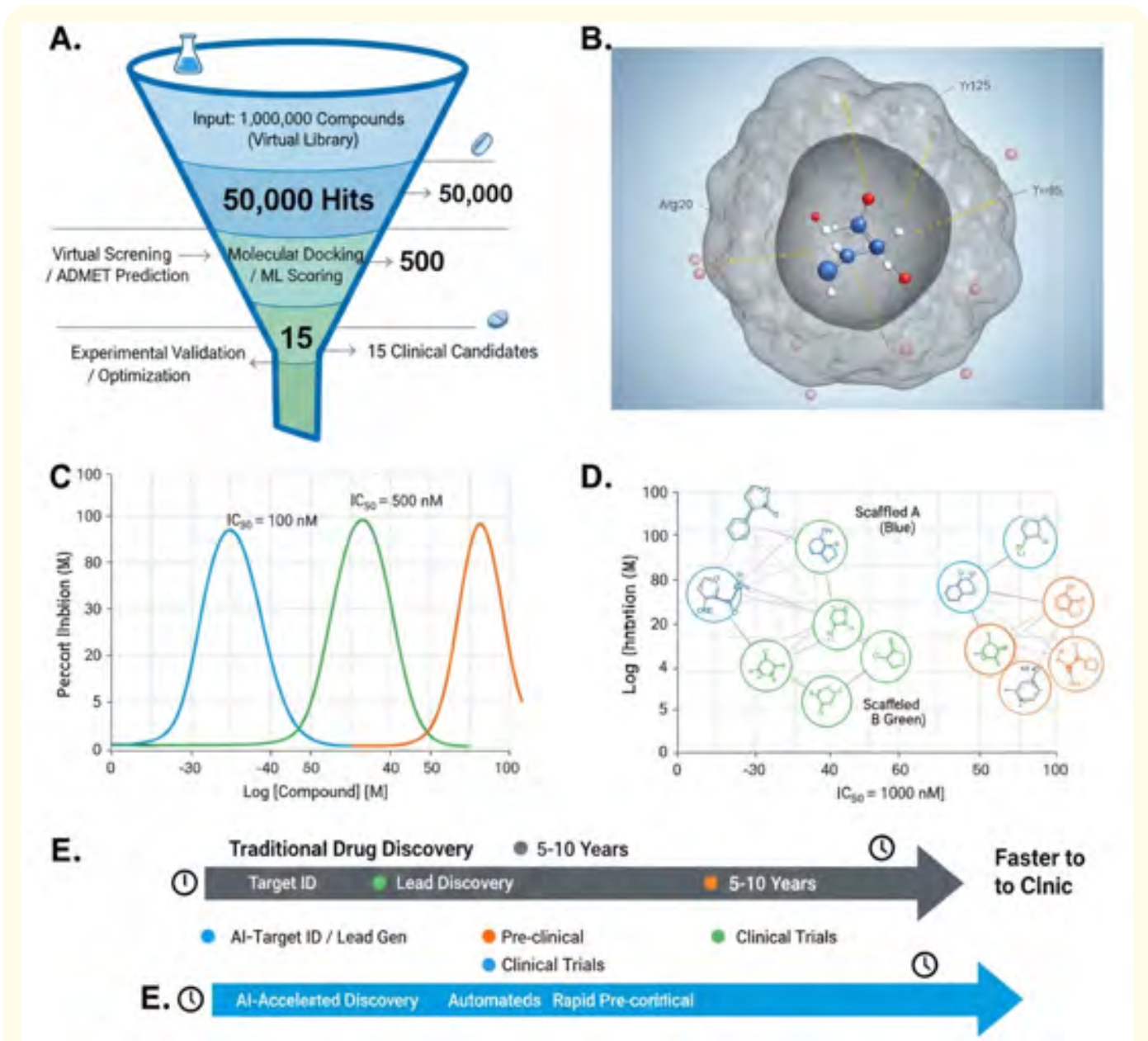
Development of explainable AI methods tailored to biological data will enhance trust and adoption of computational predictions.

Techniques that highlight key molecular features driving predictions, identify potential failure modes, and quantify uncertainty will make AI more useful for guiding experimental design and therapeutic development. Furthermore, democratisation of computational tools through user-friendly interfaces and cloud-based platforms will enable broader participation in AI-driven infectious disease research [29]. Figure 4 presents the comparative performance of machine learning models for PPI prediction, while Figure 5 shows the drug discovery applications of AI-based host-pathogen modelling.



**Figure 4:** Comparative Performance of Machine Learning Models for PPI Prediction [5,6,16,21].

(A) Receiver operating characteristic (ROC) curves comparing five ML algorithms (SVM, Random Forest, Neural Network, AlphaFold-based, Ensemble) with area under the curve (AUC) values indicated. (B) Precision-recall curves showing model performance across different confidence thresholds. (C) Confusion matrices for each algorithm display true positives, false positives, true negatives, and false negatives. (D) Feature importance ranking for the Random Forest model, showing top 15 most informative features for PPI classification. (E) Cross-validation accuracy comparison across different training set sizes, demonstrating learning curves. Error bars represent standard deviation from 10-fold cross-validation.



**Figure 5:** Drug Discovery Applications of AI-Based Host-Pathogen Modelling [4,17,18].

(A) Virtual screening workflow showing an initial compound library of 1 million molecules filtered through successive stages to identify 15 experimental candidates. (B) Binding mode visualisation of top-ranked compound (blue sticks) in the active site of pathogen target protein (gray surface), with key interacting residues labelled and hydrogen bonds shown as dashed lines. (C) Dose-response curves for the top three compounds showing  $IC_{50}$  values in the cell-based viral replication assay. (D) Structural similarity network of identified compounds revealing chemical scaffolds and structure-activity relationships. (E) Timeline comparison showing traditional drug discovery (5-10 years) versus AI-accelerated approach (6-18 months) with key milestone markers. Compound structures are represented using standard chemical notation.

## Conclusion

Computational modelling and AI-based simulation have transformed our approach to understanding host–pathogen interactions, enabling rapid prediction of molecular interactions, identification of therapeutic targets, and acceleration of drug discovery. The integration of machine learning algorithms with structural prediction, molecular dynamics simulations, and systems biology approaches provides unprecedented insights into the complex molecular choreography of infection. While significant challenges remain regarding data quality, model interpretability, and computational resources, ongoing technological advances and methodological innovations continue to expand the capabilities and impact of computational approaches.

The COVID-19 pandemic demonstrated both the potential and necessity of computational tools in responding to emerging infectious diseases. As we face continued threats from novel pathogens, antimicrobial resistance, and re-emerging infections, the synergistic application of AI with experimental biology will be essential for protecting global health. Future developments in explainable AI, multi-scale modelling, and active learning promise to further enhance our ability to predict, prevent, and treat infectious diseases. By addressing current limitations and fostering interdisciplinary collaboration, computational approaches will play an increasingly central role in infectious disease research and therapeutic development.

## Acknowledgments

The authors acknowledge the contributions of researchers worldwide whose work in computational biology and infectious disease research has advanced this field.

## Conflict of Interest

The authors declare no conflicts of interest.

## Funding

This review has no external and internal funding from any organisation.

## Bibliography

1. Kraemer MUG., *et al.* “Artificial intelligence for modelling infectious disease epidemics”. *Nature* 638.8051 (2025): 623-35.
2. Li Y and Zhou HY. “Integrating artificial intelligence with mechanistic epidemiological modeling: a scoping review of opportunities and challenges”. *Nature Communication* 16.1 (2025): 581.
3. Saluri M., *et al.* “AI-first structural identification of pathogenic protein target interfaces”. *PLOS Computational Biology* 21.6 (2025): e1013168.
4. Trepte P., *et al.* “AI-guided pipeline for protein-protein interaction drug discovery identifies a SARS-CoV-2 inhibitor”. *Molecular Systems Biology* 20.4 (2024): 428-457.
5. Singh AV., *et al.* “Machine learning and artificial intelligence for the prediction of host-pathogen interactions: a viral case”. *Frontiers in Microbiology* 12 (2021): 738459.
6. Rahmatbakhsh M., *et al.* “HPiP: an R/Bioconductor package for predicting host-pathogen protein-protein interactions from protein sequences using ensemble machine learning approach”. *Bioinformatics Advances* 2.1 (2022): vbac038.
7. Singh J., *et al.* “Machine learning for prediction of drug targets in microbe associated cardiovascular diseases by incorporating host-pathogen interaction network parameters”. *Molecular Information* 41.5-6 (2002): e2100115.
8. Kumar A and Sharma P. “Recent progress and future challenges in structure-based protein-protein interaction prediction”. *Current Opinion in Structural Biology* 92 (2025): 102771.
9. Bryant P., *et al.* “Protein interactions in human pathogens revealed through deep learning”. *Nature Microbiology* 9.9 (2024): 2370-2382.
10. Baric RS., *et al.* “A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm”. *mBio* 2.1 (2011): e00325-10.
11. Davis CN., *et al.* “Leveraging computational modeling to understand infectious diseases”. *Current Opinion on Microbiology* 56 (2020): 84-94.
12. Zhou HY and Li Y. “Unleashing the potential of artificial intelligence in infectious diseases”. *National Science Review* 12.3 (2025): nwaf004.
13. Machado MR and Pantano S. “Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses”. *Progress in Biophysics and Molecular Biology* 157 (2020): 3-13.

14. Amadei A., *et al.* “Molecular dynamics simulations: principles and applications in biology”. In: *Computational Approaches in Pharmaceutical Sciences*. CRC Press (2021): 115-44.
15. Noé F., *et al.* “Machine learning for molecular simulation”. *Annual Review of Physical Chemistry* 71 (2020): 361-390.
16. Chen Y., *et al.* “Machine learning to predict de novo protein-protein interactions”. *Trends in Biotechnology* 43.6 (2025): 772-785.
17. Schauerl M and Denny RA. “AI-based protein structure prediction in drug discovery: impacts and challenges”. *Journal of Chemical Information and Modeling* 62.13 (2022): 3142-3156.
18. Ercin N., *et al.* “Uncovering therapeutic potential through computer-based drug design”. *Computational Biology and Chemistry* 112 (2024): 108180.
19. Kortemme T. “De novo protein design: from new structures to programmable functions”. *Cell* 187.3 (2024): 526-544.
20. Dauparas J., *et al.* “Robust deep learning-based protein sequence design using Protein MPNN”. *Science* 378.6615 (2022): 49-56.
21. Ahmad Z., *et al.* “Predicting host-pathogen interactions with machine learning algorithms: a scoping review”. *Computational and Structural Biotechnology Journal* 28 (2025): 1567-1578.
22. Baruah C., *et al.* “A review of recent advances in translational bioinformatics and systems biomedicine”. In: *Information Retrieval in Bioinformatics*. Singapore: Springer (2022): 37-62.
23. Iqbal N and Kumar P. “From data science to bioscience: emerging era of bioinformatics applications, tools and challenges”. *Procedia Computer Science* 218 (2023): 1516-1528.
24. Rubin DM., *et al.* “Facilitating understanding, modeling and simulation of infectious disease epidemics in the age of COVID-19”. *Frontiers in Public Health* 9 (2021): 593417.
25. Rives A., *et al.* “Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences”. *Proceedings of the National Academy of Sciences of the United States of America* 118.15 (2021): e2016239118.
26. Gainza P., *et al.* “De novo design of protein interactions with learned surface fingerprints”. *Nature* 617.7959 (2023): 176-184.
27. Gubela N., *et al.* “Efficient and accurate simulation of infectious diseases on adaptive networks”. *PLoS Complex Systems* 1.2 (2025): e0000049.
28. Ali A., *et al.* “Dynamics of infectious disease mathematical model through unsupervised stochastic neural network paradigm”. *Computational Biology and Chemistry* 115 (2025): 108302.
29. Chalkidis G., *et al.* “Mathematical modeling of infectious disease dynamics”. *Virulence* 7.6 (2016): 636-639.