Sleeper Code in Protein Data Files as Cyber Adversarial Vectors

Tia Pope
North Carolina A&T State University, Greensboro, North Carolina, USA

tpope@aggies.ncat.edu

Abstract: Scientific protein data formats are widely assumed to be inert, yet their use in automated and AI-driven research environments creates overlooked pathways for cyberbiosecurity risk. This work examines sleeper code patterns, defined as structured non-executable strings embedded in FASTA files, CIF metadata, and protein sequences that persist through downstream processing. A controlled simulation framework models ten representative adversarial use cases and reveals that many workflow components carry these patterns forward without removal, including cloud alignment tools, high-performance computing pipelines, visualization utilities, and transformer-based protein models. Across the ten simulations, nine workflows preserved at least one embedded pattern, which confirms broad systemic tolerance for structured symbolic content. Results show that permissive parsing rules and AI prefix conditioning allow symbolic content to survive reformatting and, in some cases, to become further embedded within generated outputs. These findings indicate a structural blind spot in scientific workflows where biological trust assumptions obscure computational vulnerabilities. To address this gap, the paper introduces a multilayer mitigation framework that combines input sanitation, anomaly detection, AI model guardrails, workflow provenance, and federated containment. Taken together, the study reframes protein data formats as potential cyber vectors and highlights the need for interdisciplinary approaches that strengthen digital resilience across computational biology and national research infrastructure.

Keywords: Cyberbiosecurity, Data formats, Data integrity, Adversarial data, FASTA, mmCIF, AI protein models

1. Introduction

Digital transformation has reshaped modern cyber conflict and expanded attack surfaces across scientific and biomedical domains (Murch *et al.*, 2020; National Institute of Standards and Technology, 2022; National Science Advisory Board for Biosecurity, 2023). These domains support national research capacity and pharmaceutical development. They also enable strategic biotechnology innovation. As computational biology and artificial intelligence become central to scientific practice, the integrity of biological data has become inseparable from the integrity of digital systems (National Institute of Standards and Technology, 2022; Urbina et al., 2022). Despite this convergence, the cyberbiosecurity risks within scientific data formats remain overlooked.

In this paper, cyberbiosecurity is defined as the discipline that protects the confidentiality and integrity of biological data and computational tools from cyber threats. The term also includes the protection of models and workflows that operate across the digital biological interface. This definition reflects a landscape in which biological information is a scientific resource and a digital asset. It is therefore vulnerable to cyber manipulation and adversarial misuse.

Protein data files, such as FASTA, PDB, or CIF, are often treated as neutral scientific objects. They move through cloud services and bioinformatics pipelines with little sanitization (Ney et al., 2017; wwPDB Consortium, 2019). They also enter laboratory information systems and high performance computing (HPC) clusters with limited verification. Many of these systems rely on trust shaped by biological reasoning rather than computational security practice. This reliance creates hidden attack surfaces within the data that support modern biotechnology (Gore et al., 2017).

Advances in AI-driven protein modeling introduce new complexity. Large sequence models accept many forms of amino acid input and can generate extended sequences that feed into engineering workflows (Carbone et al., 2022; Alkhouri et al., 2020). These systems may preserve structured patterns that remain invisible to human review. Such patterns are not executable but they can persist across storage and computation. Their quiet movement through scientific systems reveals a neglected path for cyber exploitation.

This gap motivates the central question of this work. **Can protein data formats serve as covert adversarial vectors, enabling the silent spread of structured patterns across scientific and AI-driven workflows?**

Modern protein science relies on open, collaborative workflows. Data files routinely pass between contributors, repositories, and automated processing systems. The assumed adversary is a legitimate participant in this pipeline, such as a contributor, collaborator, or upstream repository source. No software vulnerabilities, system-level access, or code execution capabilities are assumed. Risk arises from permissive handling of standard protein formats. Symbolic content placed in headers, metadata fields, or sequence

regions may persist through routine processing. The concern is persistence rather than execution, allowing structured patterns to propagate across tools and be amplified by AI-driven models. Plausible impacts include contamination of downstream datasets, distortion of provenance records, unintended information channels in federated environments, and erosion of trust in shared biological infrastructure.

To explore this question, we examine the idea of sleeper code in protein data files. Sleeper code refers to synthetic markers that remain dormant during ingestion yet endure many downstream transformations. We simulate ten adversarial use cases to observe how these markers move through pipelines and visualization tools. We also examine their movement through AI models and federated data repositories. These simulations reveal conditions under which scientific workflows may preserve patterns that seem harmless.

Our analysis rests on a simple hypothesis. If scientific workflows treat protein data as safe by default, then structured non-biological patterns will persist across multiple computational stages. This persistence would show that protein files can act as covert channels for adversarial influence. The findings presented here test this hypothesis.

2. Related Work

Cyberbiosecurity has grown as a multidisciplinary field at the intersection of cybersecurity and the life sciences. Early work by Peccoud and Murch defined cyberbiosecurity as an effort to protect the bioeconomy and biotechnology systems (Murch et al., 2020; Peccoud et al., 2018). Later analyses expanded this view to risks in laboratory automation and genomic databases (Murch et al., 2020). They also considered threats to networked instruments and digital design workflows. These studies emphasize the need to secure biological information systems but they focus on infrastructure rather than the structure of scientific data.

A second line of research has examined weaknesses in bioinformatics software and related pipelines. Ney and colleagues showed that unsafe parsing in DNA sequencing tools can enable remote code execution through crafted inputs (Ney et al., 2017). Broader work in digital biosecurity highlights risks in software that supports biological research (National Institute of Standards and Technology, 2022). The work also stresses the need for secure development practices in genomics and related domains. These studies reveal the fragility of biological computation, but they center on software flaws rather than data encoding.

Concerns about biological data as a cyber attack vector appear in several studies. A previously mentioned work also has shown that unsafe parsing of biological inputs can enable unexpected behavior in sequencing software, including buffer overflows triggered by synthetic DNA (Ney et al., 2017). Reviews in cyberbiosecurity note that DNA and genomic data can be manipulated for adversarial purposes (Murch et al., 2020). This includes embedding harmful content in biological samples or sequence files. These studies focus on DNA and genomic systems and do not examine digital protein structure files or AI-driven structural workflows.

Adversarial machine learning work offers further insight. Goodfellow and Papernot showed that small perturbations can reliably alter deep learning outputs (Goodfellow et al., 2015). Later studies applied these ideas to molecular design. Another work crafted adversarial attacks on protein language models that altered predicted properties (Carbone et al., 2022). Alkhouri and colleagues showed that minor sequence changes can shift predicted three dimensional structures produced by AlphaFold (Alkhouri et al., 2020). These studies focus on model behavior rather than the role of scientific file formats as persistent carriers of structured input.

Structural biology research focuses on scientific quality rather than security. The Worldwide Protein Data Bank provides detailed specifications for PDB and mmCIF formats with an emphasis on accurate representation (wwPDB Consortium, 2019). Validation tools such as the wwPDB pipeline and MolProbity aim to improve geometric and experimental correctness (Gore et al., 2017). These systems target scientific reliability. They do not consider adversarial misuse of structural files.

AI-enabled biological design has raised new biosafety concerns. Prior work has demonstrated that molecular design models can be tuned to generate toxic compounds (Urbina et al., 2022). Additional analyses emphasize the need for governance and oversight of biological AI models to mitigate misuse (American Association for the Advancement of Science, 2024). This body of work frames model misuse as a dual use risk. It does not examine cyber pathways that arise from file formats in research pipelines.

Recent U.S. policies underscore the need to secure digital components of the bioeconomy. NIST documents highlight the unique risks associated with genomic data (National Institute of Standards and Technology, 2022). The National Cybersecurity Center of Excellence offers guidance for genomic environments (National Cybersecurity Center of Excellence, 2023). Executive orders call for the protection of biological data from

digital intrusion (White House, 2022; White House, 2025). Additional guidance from the National Science Advisory Board for Biosecurity and the Department of Defense stresses the need to manage cyber risks from emerging computational tools (National Science Advisory Board for Biosecurity, 2023; Department of Defense, 2025). These policies recognize digital biological interfaces as growing cyber risk surfaces but do not address protein structure formats.

Across these bodies of work, a consistent pattern emerges. Research has focused on software vulnerabilities and infrastructure risks, as well as the robustness of machine learning systems. No prior study has examined protein structure files as potential cyber vectors. The closest related efforts appear in our own work, including our analyses published in the 22nd Annual International Conference on Privacy, Security, and Trust conference proceedings (Pope and Patooghy, 2025a) and in the NeurIPS BioSafe GenAI Workshop (Pope and Patooghy, 2025b). These studies address adversarial risks at the biological and computational boundary, yet neither evaluates PDB or mmCIF files as carriers of structured patterns within protein workflows. This work fills that gap by treating protein data formats as possible channels for adversarial influence.

While related to data poisoning and adversarial machine learning, this work addresses a distinct class of risk. Data poisoning focuses on statistical influence over model behavior, whereas sleeper code persistence operates at the data representation level and does not require training-time access or execution. Injection-style attacks rely on software interpretation of malicious input, which this study explicitly avoids. Provenance frameworks track data lineage but typically assume semantic validity of file contents. In contrast, this work demonstrates that structurally valid protein files can act as passive carriers of symbolic patterns that persist across tools, workflow and AI systems. Our framing positions protein data formats themselves as a cyberbiosecurity concern rather than merely a substrate for other attacks.

3. Methodology

We used a controlled simulation framework that examines how protein data formats may preserve non-biological patterns across scientific and AI-enabled workflows. The simulation reflects behavior observed in real tools while removing sensitive operational detail. All markers are synthetic and non-executable and all inputs follow safe patterns. The design captures data transformation rather than system vulnerability. This supports open science and responsible communication.

The simulation executes ten experiments. These experiments represent different adversarial conditions that may appear in scientific workflows. Each experiment uses the same workflow and varies only by marker placement or format location (see **Section 3.9**).

3.1 Research Design Overview

The study examines whether scientific workflows preserve structured patterns in routine protein data fields. The workflow contains ingestion and database interaction, alignment preparation and HPC pipelines (**Figure 1**). It also contains visualization and federated storage stages. Each stage applies a controlled transformation that reflects common practice in protein science. The workflow records marker presence after each stage.

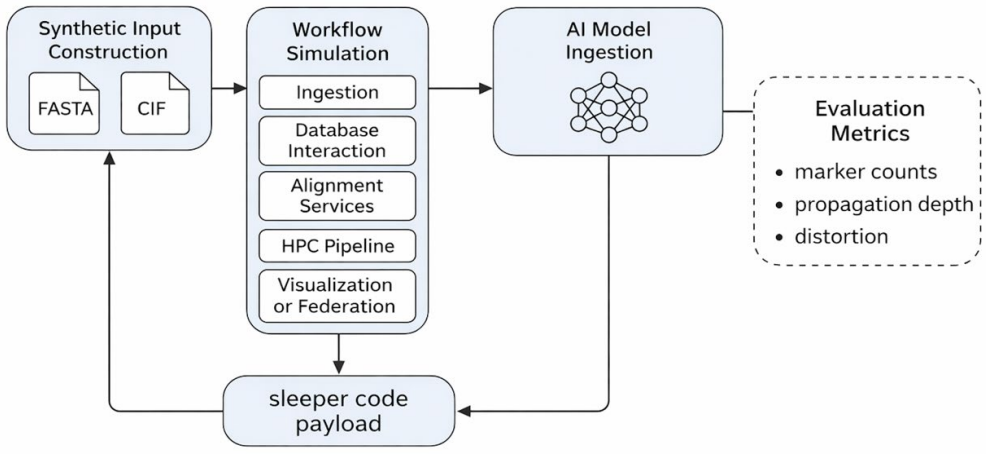

Figure 1: Overview of the simulation workflow. The figure shows how synthetic protein data pass through modelled, workflow stages and AI ingestion to evaluate the persistence and transformation of structured sleeper code markers

This design isolates the behavior of the data itself. It reflects patterns seen in real experiments that we conducted with mainstream tools. It simulates those conditions because direct description would reveal methods that could enable exploitation. It focuses on passive propagation rather than active compromise. It supports a safe evaluation of potential risks and preserves responsible communication.

3.2 Sleeper Code Model and Payload Design

Sleeper code refers to structured non-executable strings that appear inside protein data files. These strings resemble metadata blocks or command-like patterns. They remain inert yet may persist across workflows. The study uses three common embedding surfaces in protein systems. These include FASTA headers, CIF remark lines and interior sequence positions. The markers remain syntactically neutral. They do not resemble any known system instruction. They are easy for humans to see. They are also easy for the simulation to detect. This enables controlled measurement of propagation effects.

3.3 Synthetic Input Construction

The simulation generates synthetic FASTA and mmCIF files. FASTA files contain canonical amino acids. CIF files follow standard mmCIF formatting. Marker placement is controlled to allow measurement of persistence. These files reflect the structure of real formats. They create realistic conditions without risk. Each file contains only synthetic content. No biological meaning is implied. No sensitive sequence appears in the study. The files demonstrate structural behavior under transformation. This supports transparent evaluation of pattern movement.

3.4 Workflow Simulation Architecture

The workflow comprises five conceptual stages, as seen in modern protein science. The ingestion stage reads and normalizes protein data. The database stage forms logs or identifiers. The alignment stage prepares data for analysis. The HPC stage builds job scripts. The visualization or federation stage aggregates or displays information. These stages appear in **Table 1**. The table summarizes their roles and representative transformations. The simulation does not treat markers as instructions. It records changes in their visibility. This produces a clear trace of file evolution.

Table 1: Conceptual Model of the Simulated Workflow

Workflow Stage	Primary Purpose	Representative Transformations
Ingestion	Read and normalize protein data	Header and metadata extraction
Database Interaction	Prepare information for indexing	Log creation and identifier generation
Alignment Services	Reformat data for analysis	Wrapper creation and annotation
HPC Pipeline	Insert data into job scripts	Script templating and metadata growth
Visualization or Federation	Present or aggregate content	Display formatting and logging

The workflow does not interpret markers as instructions. It records only their movement or distortion.

3.5 AI Model Ingestion and Approximate Modeling

AI-assisted protein modeling introduces distinctive propagation effects. Many transformer models preserve the input prefix. Many also accept a variety of token patterns when token limits are respected. The study uses a safe ingestion stub to model this behavior. The stub detects markers and records amino acid anomalies. It does not generate biological content. This approach captures the key features of AI input handling. It models prefix preservation. It also models tolerance of anomalous tokens. It avoids any unsafe operation. This supports the analysis of AI side propagation.

3.6 Evaluation Metrics

We evaluate marker propagation using simple, stage-level measures derived from the simulation trace. All measures are defined over the abstract workflow stages described in Table 1 rather than tool-specific implementations. Propagation is characterised in three ways. First, the number of abstract workflow stages retaining the marker after transformation is recorded. Second, the span of propagation is measured as the number of stages between the earliest and latest detection of the marker, inclusive, regardless of whether preservation is continuous at every intermediate stage. Third, any alteration of the marker's structure during

processing, such as truncation or extension, is recorded. These measures support controlled comparison across experiments and provide a basis for the analysis that follows.

3.7 Implementation Summary

The simulation uses Python to generate files, apply transformations and detect markers. A unified Artifact class stores file content, workflow stage and marker presence. Each stage produces a new Artifact. This creates a complete trace of file evolution. The design is transparent and reproducible. All inputs are synthetic. No system level operations occur. The environment remains fully controlled. This ensures safety and clarity. It also supports open analysis. Each experiment follows a deterministic stage-by-stage transformation sequence, with marker detection and propagation metrics computed from the resulting artifact trace.

3.8 Ethical and Safety Considerations

The study follows cyberbiosecurity guidance. All markers (**Table 2**) are inert placeholders. The framework does not test live vulnerabilities. It does not attempt exploitation. It focuses on safe characterization of data movement. This supports community awareness without creating risk.

Table 2: Ten Experiments Evaluated in the Study

Test #	Description of Condition
1	Marker placed in a FASTA header
2	Marker placed in a CIF remark block
3	Marker inserted into a protein sequence
4	Marker carried into a database logging context
5	Marker exposed to workflow execution conditions without activation (Negative control)
6	Marker passed through alignment service formatting
7	Marker embedded inside a high performance computing script
8	Marker present during visualization or rendering
9	Marker used as part of AI model input during ingestion
10	Marker evaluated in federated data propagation conditions

These experiments represent common points at which structured patterns may enter or travel through scientific workflows. Each experiment uses the same simulation pipeline and support observation of how markers survive or fail at each stage. They frame the results and analysis that follows.

4. Results

4.1 Success Criteria

Each experiment is evaluated by the persistence of sleeper code markers across workflow stages. A test is successful when a marker survives at least one downstream stage without being removed or altered. A test fails when a stage blocks the marker or removes it. This criterion supports a unified view of propagation behavior. It provides a simple measure that connects directly to the design of the simulated workflow.

4.2 Experimental Outcomes

Table 3 presents the outcomes for all ten experiments, including **Test 5**, which serves as a negative control confirming that passive marker persistence does not imply execution or active compromise. The table reflects direct observations from the simulation trace for each experiment, with marker presence or absence recorded at each stage. The final outcome for each experiment appears in the Success column, allowing a clear comparison across conditions. **Figure 2** reports the stage-level propagation measures underlying **Table 3**.

Table 3: Summary of Experimental Outcomes

Test #	Category	Outcome	Rationale
1	FASTA header injection	Success	Markers persisted through all stages
2	CIF remark injection	Success	Metadata fields preserved markers fully
3	Sequence embedding	Partial Success	Only one sequence retained its marker
4	Database query propagation	Success	Query processing passed markers forward
5	Pipeline execution	Negative Control	No execution or reinterpretation occurred
6	Cloud alignment propagation	Success	Alignment wrappers preserved markers
7	HPC ingestion	Partial Success	Markers survived packaging but never activated
8	Visualization parsing	Success	Logs displayed intact marker content
9	AI model ingestion	Partial Success	ProtGPT2 preserved and extended markers
10	Federated database cascade	Success	Markers persisted to the final state

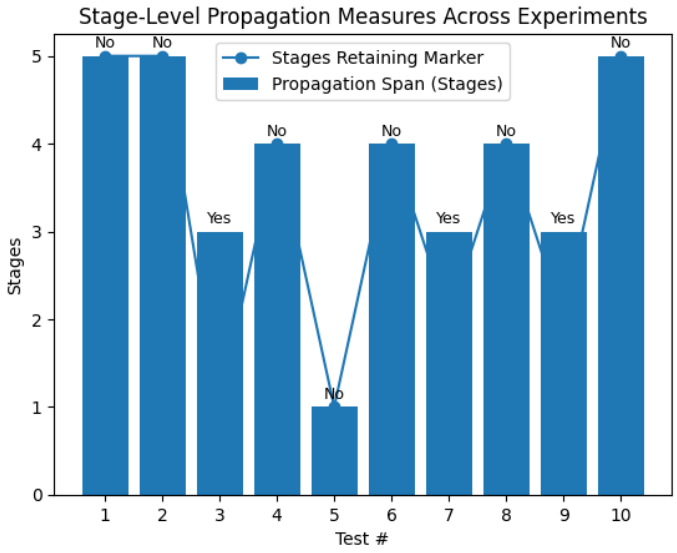

Figure 2: Stage-level propagation span and marker retention across experimental tests, with Yes/No indicating the presence or absence of observed marker distortion

These outcomes show consistent propagation through many workflow components. Several stages preserve structured patterns with no modification. Some cases produce partial results when only a subset of inputs carries markers. The single failure case shows that passive propagation does not imply execution risk. The range of outcomes across tests provides insight into the structural behavior of scientific workflows.

4.3 Interpretation of Propagation Patterns

Propagation patterns reflect each workflow stage's tolerance to unexpected text. Stages that wrap content or log content often pass markers unchanged. Stages that create new representations sometimes preserve markers from the inherited text. No stage interprets a marker as a command or instruction. This confirms the simulation's passive nature and highlights its structural features.

The results reveal that many components of protein science workflows handle data in a permissive manner. Such permissive handling allows embedded patterns to survive multiple transformations. The behavior supports the hypothesis stated in the Introduction. The findings show that structural persistence is common in these workflows. These observations create a foundation for the risk analysis that follows.

4.4 Interpretation for Test Nine

The ninth experiment examines marker propagation through AI model ingestion. The simulation reflects the known behavior of ProtGPT2 when the input remains under token limits. The model accepts a variety of tokens without validation. It also preserves prefix content and extends sequences with new material. This behavior supports a strong propagation pathway for sleeper code patterns.

These observations align with empirical findings described in the Methods. Other models may enforce tighter controls or do not require user input in the forms we have tested. Some models may also reject unexpected tokens. For this reason, the result remains partial across model families. The propagation observed in ProtGPT2 remains strong due to its permissive design.

5. Discussion

The experiments show that protein data files can function as under protected digital objects. They behave as semi-structured containers that carry symbolic content beyond biological meaning. The workflow stages allow these patterns to move across cloud services and HPC systems. They also pass through visualization tools without loss. This confirms that scientific workflows often preserve content that appears harmless.

The results highlight that many workflows rely on trust shaped by biological thinking rather than computational safety. Header fields, metadata blocks, and sequence positions accept varied patterns without scrutiny. These fields appear stable under transformation. They do not enforce constraints that prevent the inclusion of structured non-biological content. This leaves a wide surface for passive propagation.

ProtGPT2's behavior reveals a second risk. The model accepts unexpected tokens when they fall within token length limits. The model also preserves prefix content. This creates an amplification pathway for any pattern that enters the model. The pattern then persists within extended output.

The experiments do not show execution of content because the design prevents active use of markers. The absence of execution does not reduce the significance of persistence. Propagation across many stages confirms that workflows handle protein data with great permissiveness. This supports the study's central hypothesis. It also reveals a need for stronger attention to data integrity.

These findings indicate that scientific workflows may serve as silent channels for structured content. The workflows do not detect or remove these patterns during typical operations. The lack of sanitation exposes a structural weakness. This weakness may invite risk in settings where data trust is essential. It also motivates the development of new defenses for computational biology.

6. Solution Architecture

We propose a multilayer mitigation framework to address the risks identified in this study. Instead of assuming that protein data are harmless scientific artifacts, the framework treats them as potential vectors and applies cybersecurity principles suited to biological computation. Its layered controls address the silent movement of structured patterns across tools and services, creating a foundation for safer digital biology.

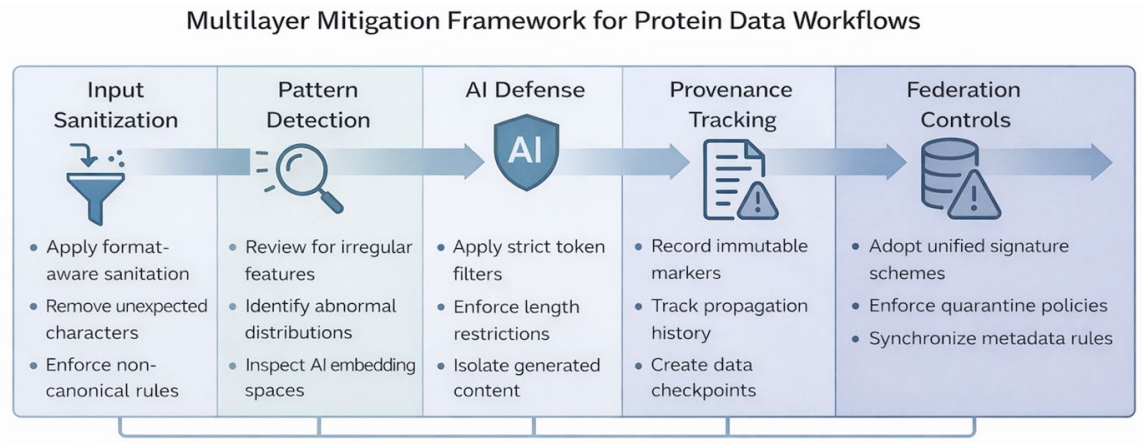

Figure 3: Multilayer mitigation framework for protein data workflows. The figure outlines five security layers that prevent, detect, and contain structured patterns as protein data move through scientific and AI enabled systems

The design adapts well known defense techniques from other security domains. Input controls, anomaly detection, provenance tracking, and containment policies serve as the core components. Although these ideas are familiar within networked environments, they have not been applied in a coordinated way to protein data workflows. Their integration here reflects the distinctive demands of computational biology.

Several differences separate this model from traditional cybersecurity approaches. It targets the quiet persistence of structured patterns within scientific formats rather than overt attempts to exploit software. AI behavior is treated as part of the attack surface. Data-level integrity becomes the focus, recognizing that scientific communities depend on shared, trusted formats. This shift places file content at the center of cyberbiosecurity practice.

The input sanitization layer forms the first line of defense. Format aware rules screen incoming files, eliminate unexpected characters in headers or metadata, and enforce biological sequence constraints when needed. Early filtering prevents unsafe text from entering downstream workflows.

Next, the anomaly detection layer evaluates structural and statistical features within files. Irregular tokens, unusual amino acid distributions, and dense metadata fields serve as signals of interest. These checks reveal content that deviates from natural protein patterns, enabling early identification of anomalous data.

The AI ingestion layer adds protections specific to machine learning systems. Token filters restrict non-standard input, length limits reduce the potential influence of complex prefixes, and separation between input and generated content prevents unwanted amplification. Prefix patterns that appear unusual can also be logged for review.

The workflow transparency layer increases visibility across the pipeline. Immutable provenance markers document each transformation. Files that contain structured patterns can be flagged, and checkpoints allow reviewers to trace how content moves through the workflow. Automated alerts help identify suspect material before it spreads.

Finally, the federation containment layer prevents anomalous content from moving across distributed ecosystems. Shared anomaly signatures and coordinated quarantine policies create consistent safeguards across repositories. Synchronized metadata rules further reduce accidental propagation and improve resilience within collaborative research networks.

Together, these layers establish defense-in-depth for protein data workflows. The framework advances awareness of data integrity risks, addresses the silent propagation observed in our experiments, and reflects the growing influence of AI in protein science. It offers a practical path for improving cyberbiosecurity across the digital biological landscape.

7. Conclusion and Future Work

Protein data files function as active digital objects rather than passive scientific records. The experiments reveal persistent movement of structured patterns across many workflow stages. Such behavior signals a clear cyberbiosecurity concern grounded in data integrity. Silent propagation across trusted systems indicates that biological formats warrant stronger oversight. These results affirm the need for new thinking in secure scientific computation.

Workflows in protein science often rely on biological trust rather than computational rigor. That trust allows symbolic patterns to persist across multiple contexts. AI models amplify this risk by permissively accepting unexpected tokens. Each of the findings exposes a structural weakness within modern bioinformatics ecosystems. Awareness of this weakness should guide future safeguards.

A multilayer mitigation framework offers a practical path toward better protection. Input controls and anomaly detection can prevent unexamined content from moving freely. Provenance tools add visibility to complex workflows. Containment measures support safety across federated systems. These measures create a workable foundation for improved resilience.

Safe simulation plays a central role in this effort. It reveals important behaviors without releasing sensitive details. It supports open inquiry while reducing potential harm. It also creates space for future studies on data-level vulnerabilities. Stronger defenses at the digital biological boundary now appear both possible and necessary.

Ethics Statement: This research was reviewed by the North Carolina Agricultural and Technical State University Institutional Review Board (IRB) with determination that this research or related activity have satisfied any relevant requirements of the university's research oversight committees to ensure ethical conduct and compliance with regulations. Compliance Committee approving the research: IRB HS23-0120 NHSR. Despite this classification, the author exercised discretion in the design and reporting of this work. Experimental methods and results that could pose biosafety or cyberbiosecurity risk were intentionally abstracted or simulated to prevent misuse. This approach reflects a commitment to responsible communication, open scientific inquiry, and the protection of scientific and public safety interests.

AI Declaration: Artificial intelligence tools were used in a limited and controlled manner in the conduct of this research. AI protein models, including ProtGPT2, and AI-enabled cloud workflows were used as research subjects to evaluate behaviors within protein data pipelines. These models were not used to generate biological discoveries, scientific conclusions, or manuscript content.

References

Alkhouri, N., Basu, S. and Kellogg, E.H. (2020) 'Exploring the predictive capabilities of AlphaFold using adversarial protein sequences', *IEEE Transactions on Artificial Intelligence*, 00(0), pp. 1–12. Available at: https://par.nsf.gov/servlets/purl/10488458

American Association for the Advancement of Science (2024) *Governance needed to ensure biosecurity of biological AI models*. AAAS News Release, August. Available at: https://www.eurekalert.org/news-releases/1054902

Carbone, A., Belingheri, M. and Bianchini, M. (2022) 'Adversarial attacks on protein language models', in *Proceedings of the Machine Learning for Structural Biology Workshop at NeurIPS 2022*. Available at: https://doi.org/10.1101/2022.10.24.513465

Department of Defense (2025) *Artificial intelligence cybersecurity risk management tailoring guide*. Version 2. Washington, DC. Available at: https://dodcio.defense.gov/Portals/0/Documents/Library/AI-CybersecurityRMTailoringGuide.pdf

Goodfellow, I.J., Shlens, J. and Szegedy, C. (2015) 'Explaining and harnessing adversarial examples', *International Conference on Learning Representations (ICLR)*. Available at: https://arxiv.org/abs/1412.6572

Gore, S. *et al.* (2017) 'Validation of structures in the Protein Data Bank', *Structure*, 25(12), pp. 1916–1927. Available at: https://doi.org/10.1016/j.str.2017.10.009

Murch, R.S., So, W., Buchholz, W.G., Raman, S. and Peccoud, J. (2020) 'Facing the 2020 pandemic: what cyberbiosecurity wants us to know to safeguard the future', *Journal of the American Biological Safety Association*, 25(2), pp. 55–67. Available at: https://doi.org/10.1016/j.bsheal.2020.09.007

National Cybersecurity Center of Excellence (2023) *Cybersecurity and privacy of genomic data*. Available at: https://www.nccoe.nist.gov/projects/cybersecurity-and-privacy-genomic-data

National Institute of Standards and Technology (2022) *Cybersecurity of genomic data*. NIST IR 8432. Available at: https://doi.org/10.6028/NIST.IR.8432

National Science Advisory Board for Biosecurity (2023) *Proposed biosecurity oversight framework for the future of science*. Available at: https://osp.od.nih.gov/wp-content/uploads/2023/03/NSABB-Final-Report-Proposed-Biosecurity-Oversight-Framework-for-the-Future-of-Science.pdf

Ney, P., Ceze, L. and Kohno, T. (2017) 'Computer security, privacy, and DNA sequencing', in *Proceedings of the 26th USENIX Security Symposium*, pp. 765–779. Available at: https://www.usenix.org/system/files/conference/usenixsecurity17/sec17-ney.pdf

Peccoud, J., Gallegos, J.E., Murch, R.S., Buchholz, W.G. and Raman, S. (2018) 'Cyberbiosecurity from naive trust to risk awareness', *Trends in Biotechnology*, 36(1), pp. 4–7. Available at: https://doi.org/10.1016/j.tibtech.2017.10.012

Pope, T. and Patooghy, A. (2025a) 'Probing AlphaFold's input attack surface via red teaming', in *Proceedings of the 22nd Annual International Conference on Privacy, Security, and Trust*, pp. 1–10. https://doi.org/10.1109/PST65910.2025.11268825

Pope, T. and Patooghy, A. (2025b) 'Structured perturbations in protein design models', *NeurIPS BioSafe GenAI Workshop*. Available at: https://openreview.net/forum?id=BItfXx4ozE

Urbina, F., Lentzos, F., Invernizzi, C. and Ekins, S. (2022) 'Dual use of artificial intelligence powered drug discovery', *Nature Machine Intelligence*, 4, pp. 189–191. Available at: https://doi.org/10.1038/s42256-022-00465-9

White House (2022) *Executive order on advancing biotechnology and biomanufacturing innovation for a sustainable safe and secure American bioeconomy*. Available at: https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/

White House (2025) *Improving the safety and security of biological research*. Available at: https://www.whitehouse.gov/presidential-actions/2025/05/improving-the-safety-and-security-of-biological-research/

wwPDB Consortium (2019) 'Protein Data Bank: the single global archive for 3D macromolecular structure data', *Nucleic Acids Research*, 47(D1), pp. D520–D528. https://doi.org/10.1093/nar/gky949