AI Bioweapons and the Failure of Inference-Time Filters

Jonathan Feldman
Georgia Institute of Technology
Atlanta, GA
jonathanfeldman@gatech.edu

Tal Feldman
Yale Law School
New Haven, CT
tal.feldman@yale.edu

Abstract

Recent advances in generative biology have enabled the design of novel proteins, creating significant opportunities for drug discovery while also introducing new risks, including the potential development of synthetic bioweapons. Existing biosafety measures primarily rely on inference-time filters such as sequence alignment and protein–protein interaction (PPI) prediction to detect dangerous outputs. In this study, we evaluate the performance of three leading PPI prediction tools: AlphaFold 3, AF3Complex, and SpatialPPIv2. These models were tested on well-characterized viral–host interactions, such as those involving Hepatitis B and SARS-CoV-2. Despite being trained on many of the same viruses, the models fail to detect a substantial number of known interactions. Strikingly, none of the tools successfully identify any of the four experimentally validated SARS-CoV-2 mutants with confirmed binding. These findings suggest that current predictive filters are inadequate for reliably flagging even known biological threats and are even more unlikely to detect novel ones. We argue for a shift toward response-oriented infrastructure, including rapid experimental validation, adaptable biomanufacturing, and regulatory frameworks capable of operating at the speed of AI-driven developments.

1 Introduction

In recent years, the convergence of generative AI and molecular biology has produced powerful new systems capable of designing proteins with remarkable precision. Protein language models (PLMs), along with structural predictors and design tools like RFdiffusion, are revolutionizing medicine by enabling rapid therapeutic development—from novel enzymes and vaccines to antibody candidates tailored for specific diseases [1–5]. These systems are accelerating biomedical innovation, shrinking experimental timelines, and expanding the reach of drug design to previously intractable targets [6, 7].

Yet this same capability introduces an unprecedented dual-use dilemma. The tools now used to craft life-saving biologics can, in the wrong hands, be repurposed to generate pathogens with enhanced infectivity, immune evasion, or toxicity [8]. PLMs can simulate the evolutionary trajectory of viral proteins, suggesting mutations that make them more transmissible or harder to neutralize [9–11]. Structure-aware design tools can validate these outputs in silico and propose new scaffolds to support them [6, 12]. This design-and-validation loop—which mirrors workflows already used in therapeutic biotech—lowers the barrier to creating synthetic threats, some of which may be novel enough to evade detection altogether [8].

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: BioSafe GenAI.

In response to this growing concern, the AI and biosecurity communities have proposed technical countermeasures [13]. These include inference-time filtering to block dangerous outputs, red-teaming to test models for misuse, and sequence screening to flag potentially pathogenic molecules [8, 13, 14]. However, these methods rest on a problematic assumption: that we can reliably identify harmful proteins computationally.

This assumption breaks down under scrutiny. Whether a protein is benign or harmful depends on factors that remain opaque to even the most advanced models—its tertiary structure, cellular context, binding specificity, and immunogenicity [8, 14, 15]. Most of these properties cannot be inferred from sequence alone. In fact, as we demonstrate in this paper's technical section, existing models for protein–protein interaction (PPI) prediction fail especially badly when applied to viral proteins. They lack the data diversity, contextual understanding, and biological realism to generalize beyond narrow, overrepresented domains. This limitation creates blind spots that could be exploited by malicious actors or simply overlooked in good-faith deployments.

If computational filters are not fully reliable then only experiments can provide definitive safety assurance. Yet experimental validation methods, such as binding assays, toxicity screens, and infection models, are too slow, costly, and resource-intensive to scale with model outputs [16, 17]. In theory, flagged sequences could be tested in the lab. In practice, we lack the infrastructure to vet more than a tiny fraction [8, 18]. That leaves us with a dangerous asymmetry: generative models can produce threats at scale, but our ability to verify them lags far behind.

Worse, even dramatic advances in our ability to predict pathogenicity from sequence data will not guarantee security. Open-source models, fine-tuned on public data, are available worldwide. The barriers to misuse are falling: DNA synthesis services are increasingly accessible, foundational laboratory tools are widely distributed, and the necessary expertise is no longer confined to state-sponsored programs [8, 12, 19–21]. Nation-states and non-state actors with sufficient resources can run models privately, circumvent safety mechanisms, and synthesize outputs with little oversight. In this threat landscape, technical success in sequence classification does not guarantee security. We cannot rely on preemptive filters alone.

The implication is stark: current defense strategies—focused on prediction, prevention, and perimeter control—are structurally mismatched to the nature of the threat. No matter how refined our red-teaming becomes, no matter how many inference-time filters we deploy, we cannot guarantee safety when our core tools lack ground truth and when potential adversaries are unconstrained.

What is needed instead is a shift: from a paradigm of containment to one of resilience. Rather than attempting to intercept every potentially dangerous output at generation time, we must build systems capable of responding to threats after they emerge—faster than those threats can spread. This means developing AI-native infrastructure for therapeutic response: models trained to design safe, stable, and manufacturable countermeasures on demand; high-throughput simulation and validation tools that can screen thousands of candidates in parallel; and a nationwide biomanufacturing pipeline that can pivot from peacetime research to emergency deployment within days.

Crucially, this is not merely a technical aspiration—it is a national security imperative. We must prepare for a world in which synthetic pathogens may be generated and released faster than traditional public health systems can detect or contain them. In such a world, resilience is the only viable defense. Speed—not restraint—becomes the battlefield.

This paper is structured as follows: In the next section, we show that current machine learning models for predicting viral protein interactions, especially those relying on large-scale sequence and structure modeling, fail to generalize well. This is particularly evident in low-data and high-divergence scenarios. These shortcomings highlight the limitations of current safety frameworks and underscore the urgent need to shift research focus towards response-driven capabilities. We then propose a resilience-based framework that addresses gaps in existing biosecurity efforts. Ultimately, our goal is not to eliminate generative biology, but to guide it—ensuring that the tools being developed today are aligned with the future needs that will arise.

2 Limitations of Current Inference-Time Filtering Approaches

The task of identifying potentially pathogenic proteins, especially those that may be synthetically engineered or genetically modified, for inference-time filtering requires robust, reliable tools. In the

context of biosecurity, these tools must be able to predict whether a given protein sequence or structure could be harmful to humans. Currently, the two primary methodologies used to predict pathogenicity are sequence similarity comparisons and PPI prediction mechanisms [22, 14, 23, 24]. While these tools have proven useful, both are fundamentally limited in their applicability and accuracy, especially when it comes to detecting novel or engineered proteins.

Sequence similarity searches are arguably the most straightforward and effective tool for identifying known pathogenic proteins. By comparing an input protein sequence against a large database of known sequences, researchers can rapidly identify homologous proteins from established viral strains [25]. This method is particularly useful for identifying pathogenic proteins that have undergone minor mutations or are derivations of previously known viral proteins. Tools such as BLAST facilitate the comparison of protein sequences to databases such as UniProt and NCBI Virus, which contain thousands of viral sequences [26–28]. These comparisons allow for the determination of the degree of homology between the input protein and known viral sequences, providing a relatively simple and highly scalable means of detecting pathogenicity [26, 29]. When the degree of homology exceeds a pre-defined threshold, the protein in question can be flagged as potentially pathogenic.

Despite its effectiveness, sequence similarity searches are fundamentally limited by their reliance on existing databases. Since these databases are composed of known sequences, they inherently fail to detect novel viral proteins that are sufficiently divergent from anything previously cataloged [14]. This means that while sequence similarity searches can effectively identify variants of known pathogens, they are incapable of detecting entirely novel proteins, such as those designed using advanced computational techniques like RFdiffusion or ESM3, which generate novel protein sequences that are not present in any existing databases [6, 12]. Therefore, while sequence similarity searches remain a valuable tool for detecting known pathogens, they cannot be relied upon to identify entirely new threats, especially those that have been deliberately engineered.

To address this limitation, researchers have turned to PPI prediction methodologies [30, 31]. PPIs are central to the pathogenicity of many viruses, as viral proteins typically need to interact with human cellular receptors to gain entry into the host [32–34]. While most viruses bind to standard receptors, some also interact with glycoproteins and glycolipids on the surface of host cells [34]. PPI prediction tools have been developed with the aim of identifying these interactions, thereby providing a complementary approach to sequence similarity searches. In theory, PPI prediction tools could be used to determine whether a viral protein is likely to bind to a receptor on a human cell, serving as a reasonable proxy for pathogenicity. This would allow for the identification of proteins that might be capable of interacting with human receptors, even if they are not closely related to any known pathogen.

However, in practice, PPI prediction tools—especially those based on artificial intelligence—often fail to deliver the accuracy required for high-stakes applications such as biosecurity [35]. The inherent limitations of these models are particularly concerning when one considers the potential consequences of inaccuracies. A false negative, in which a dangerous pathogenic protein goes undetected, could have dire consequences for public health. For these reasons, the accuracy of PPI prediction models is a critical issue that must be addressed before these tools can be used reliably for biosecurity purposes.

In order to evaluate the performance of state-of-the-art PPI prediction models, we conducted a benchmarking study using thirteen well-characterized viral protein–human receptor interactions. These interactions were selected from viruses that have fully or partially resolved structures available in the Protein Data Bank (PDB), such as HIV, Measles, and SARS-CoV-2 [36]. Three widely used PPI prediction tools were selected for testing: AlphaFold 3, AF3Complex, and SpatialPPIv2. AlphaFold 3, developed by Google DeepMind, is a deep learning-based model that is widely regarded as the best available tool for predicting protein structures [37]. AF3Complex is a derivative of AlphaFold 3 that has been specifically designed to improve PPI predictions by modeling those interactions more accurately [38]. SpatialPPIv2 is a graph neural network-based model that represents the latest advancement in PPI prediction methodologies [23].

Each of the models was provided with the protein sequences of the viral binding protein and its corresponding human receptor. AlphaFold 3 and AF3Complex used sequence-based inputs to predict the structures of the individual proteins, while SpatialPPIv2 was provided with pre-computed structures generated by AlphaFold 3 [23, 37, 38].SpatialPPIv2 differs from AlphaFold 3 and AF3Complex in that it is explicitly trained to predict PPIs using structural inputs rather than sequence information [23]. AlphaFold 3 relies on its internally generated interface predicted

TM-score (ipTM) to evaluate the confidence of predicted protein complex structures; prior studies have used an ipTM score of 0.6 or higher as a proxy for a meaningful PPI [37, 39]. In contrast, AF3Complex uses a separate internal metric, the predicted Interface-Similarity score (pIS), with scores of 0.38 or above indicating a likely PPI [38, 40, 41]. SpatialPPIv2 instead outputs a probability that a meaningful interaction exists between the input proteins, with a threshold of 0.5 or higher used to indicate a predicted PPI.

Virus	**SpatialPPIv2**	**AlphaFold3**	**AF3Complex**
Adenovirus	✓	✓	✓
Chikungunya	×	×	×
Ebola	✓	×	✓
Epstein-Barr	✓	✓	✓
Hepatitis B	×	×	×
HIV	✓	✓	✓
HSV (HVEM Receptor)	✓	×	×
HSV (Nectin-1 Receptor)	✓	✓	✓
Measles	×	×	×
Measles (Edmonston strain)	×	✓	✓
MERS	✓	✓	✓
Nipah	✓	✓	✓
SARS-CoV-2	×	×	✓

Table 1: Virus-host protein–protein interaction prediction performance across three computational models. Each row represents a different virus strain or receptor variant, with checkmarks (✓) indicating successful PPI identification and crosses (×) indicating failed predictions. Confidence thresholds differ by model: SpatialPPIv2 (0.5), AlphaFold3 (0.6), and AF3Complex (0.38).

The results of the benchmarking study, summarized in Table 1, reveal a significant failure of all three PPI prediction models to consistently identify well-documented viral–host interactions. Notably, both AlphaFold 3 and SpatialPPIv2 failed to correctly identify the interaction between the SARS-CoV-2 spike protein and the human ACE2 receptor—an interaction that is critical to the virus's ability to infect human cells and that was at the heart of the COVID-19 pandemic [9, 42]. In total, SpatialPPIv2 misidentified approximately 40% of the PPIs, while AlphaFold 3 misidentified about 50% and AF3Complex misidentified about 30%. These results indicate that even state-of-the-art models are far from reliable when it comes to predicting viral–host interactions, which undermines the utility of these models for identifying computationally designed pathogenic proteins. Moreover, many of the PPIs included in the benchmarking study were present in the training datasets of all three models [37, 38, 43]. This means that these models had already been exposed to these interactions during training, yet they still failed to accurately predict them in the testing phase. This is a particularly troubling finding, as it suggests that the models are not generalizing well to real-world PPI predictions, even when the interactions are well-known and well-characterized. If these models cannot even accurately predict the most established viral PPIs, the possibility of them detecting novel or engineered viral–host interactions becomes even more unlikely.

As an ancillary experiment, we evaluated four mutated variants of the SARS-CoV-2 spike receptor-binding domain, the viral protein responsible for mediating COVID-19 infections. These mutations were sourced from a prior study [44], which experimentally quantified their effects on binding affinity between the spike RBD and the human ACE2 receptor, a key determinant of protein–protein interaction strength. Among the selected variants, two exhibited the greatest enhancement in ACE2 binding affinity, with more than tenfold increases, while the remaining two showed the most substantial reductions in affinity that still preserved a functional interaction.

We assessed these four spike mutants using the same three structure-based models evaluated previously—SpatialPPIv2, AlphaFold3, and AF3Complex—to determine whether any PPIs between spike and ACE2 were computationally predicted. The results are shown in Table 2.

Strikingly, all three models failed to detect meaningful interactions for any of the variants, including those with experimentally confirmed high-affinity binding. This reveals a strong insensitivity to combinatorial mutational effects, highlighting a significant blind spot: in the absence of sequence similarity searches, these models may be unable to identify pathogenic viral variants. This raises serious concerns about their utility in validating viral pathogenicity.

Even more concerning is the direction of the models' confidence shifts. For both beneficial and detrimental mutations, SpatialPPIv2 and AlphaFold3 increased their predictive confidence, while AF3Complex decreased its own, regardless of the true biological impact. This erratic behavior underscores the models' limitations when faced with mutationally altered viral proteins—a critical vector through which more virulent or evasive pathogens could emerge [8].

A major contributor to these failures is a fundamental data deficiency, a lack of high-quality, well-annotated examples of viral–host protein interactions in existing training datasets [45, 46]. Although resources like the PDB contain hundreds of thousands of protein structures, the vast majority of these represent non-viral proteins, with viral proteins, especially those relevant to host interactions, comprising only a small fraction [36, 46]. This lack of targeted data limits the models' ability to learn interaction-relevant features specific to viral proteins.

However, this data imbalance is only part of the challenge. Generalizing to novel or engineered proteins remains inherently difficult because of the complex biophysical nature of protein–protein interactions [45]. Even with improved viral datasets, predicting interactions involving synthetic or previously unseen proteins, such as those generated by modern protein design tools, presents a major obstacle. These artificial proteins may not resemble known viral proteins and may interact with host receptors in unconventional ways, making them especially difficult for models trained on natural proteins to detect [6, 12].

Affinity Change ($\Delta \log_{10} K_D$)	**SpatialPPIv2**	**AlphaFold3**	**AF3Complex**
-1.04	0.27 (+0.08)	0.16 (+0.03)	0.11 (–0.47)
$+2.14$	0.19 (+0.00)	0.24 (+0.11)	0.06 (–0.52)
-1.03	0.26 (+0.07)	0.14 (+0.01)	0.06 (–0.52)
$+2.11$	0.24 (+0.05)	0.15 (+0.02)	0.05 (–0.53)

Table 2: Performance comparison of PPI prediction models on SARS-CoV-2 spike protein mutants. Each row represents a different mutational variant with its corresponding change in binding affinity ($\Delta \log_{10} K_D$). Negative values indicate an increase in binding affinity, while positive values indicate the opposite. Parentheses show the difference from the baseline confidence score for the wildtype SARS-CoV-2 spike protein and human ACE2 PPI. SpatialPPIv2's PPI threshold is 0.5, AlphaFold3's is 0.6, and AF3Complex's is 0.38. *None of the models correctly predicted a single PPI.*

This gap underscores a broader limitation in current PPI prediction approaches. While models can perform reasonably well when applied to familiar protein types, they falter in high-stakes scenarios where generalization is critical, such as detecting engineered proteins with potential for harm. As generative tools become more accessible and capable, the inability to reliably assess the risk of novel proteins presents a serious and growing biosecurity concern.

3 Reorienting Research Toward Infrastructure Readiness and Rapid Response

Current inference-time filters and protein–protein interaction predictors aren't enough to catch dangerous synthetic proteins. As generative biological models improve and spread, it is increasingly clear that biosecurity cannot rely on static safeguards alone. Instead, the focus must shift toward building a resilient infrastructure that is capable of designing, validating, manufacturing, and deploying medical

countermeasures at machine timescales. That means leadership from federal agencies but also a broad realignment of research priorities across universities, research labs, and industry.

To date, most investment in generative AI for therapeutic design has been concentrated in the pharmaceutical industry and biotechnology startups [47, 48]. These efforts naturally focus on common diseases with strong commercial incentives and well-known biological targets. But that leaves a gap: there's little work on the rare, fast-evolving, or engineered pathogens most relevant to biodefense [8, 49, 50]. Academic research must fill this gap. Universities and consortia are well suited to study the edge cases that industry R&D tends to overlook. By targeting novel viruses, unknown protein structures, and synthetic sequences outside today's biological data, they can help build tools that matter for alignment and security.

These efforts will fail without addressing a deeper problem: the critical data needed to train and evaluate countermeasure models is scattered and siloed. Information on pathogen evolution, immune response, failed therapeutics, and synthetic constructs sits across disconnected academic, corporate, and government systems [51]. Integrating that data is essential—but so is access control. In a national security context, broad sharing is risky. Access should be limited to certified research labs and trusted partners operating under clear safeguards. We need sustained investment in shared infrastructure to support this: repositories of known and hypothetical viral strains, standardized measurements of immunogenicity and toxicity, and transparent clinical performance data. Funding agencies should prioritize these efforts, and academic institutions must be willing to build and maintain secure platforms that enable collaboration without compromising biosecurity. The COVID-19 response showed what rapid, coordinated data sharing can unlock [52]; biodefense now requires a version built for both speed and control.

Beyond data and modeling, the academic community has a responsibility to reframe how generative protein design is conceptualized. Much of the discourse surrounding PLMs and other generative tools has focused on speculative risks, particularly their potential misuse for creating harmful biological agents [8, 13, 49, 50]. But the risks posed by generative protein models are no longer theoretical, and the research community needs to treat them accordingly. As these systems grow more capable, especially in designing novel or engineered proteins, their potential for misuse becomes a concrete concern. Addressing this requires more than superficial safeguards or ad hoc filtering. It demands structural solutions. Treating biosecurity as a secondary or post hoc issue is no longer tenable. These challenges sit at the center of responsible model development and must be integrated into core research agendas.

Still, while academia can help define the technical foundations of preparedness, the responsibility for real-world deployment rests firmly with the federal government. Wet-lab validation, clinical testing, and large-scale manufacturing of AI-generated therapeutics will require coordination, supply chains resilience, and emergency infrastructure that are beyond the capacity of any individual lab or institution. Agencies such as the Biomedical Advanced Research and Development Authority (BARDA), the Food and Drug Administration (FDA), and the Administration for Strategic Preparedness and Response (ASPR) must take the lead in establishing national infrastructure that translates model output into deployable countermeasures.

This includes building experimental screening platforms that can test thousands of protein designs in parallel for binding, toxicity, manufacturability, and immune response. These systems should be fast, modular, and automated to keep pace with rapid model outputs. But screening alone isn't enough. The federal government should also support a distributed network of biomanufacturing sites, ready to produce countermeasures in a crisis. These facilities should be maintained on standby, stocked with key materials, and able to shift quickly from routine work to emergency response. Redundancy and geographic spread will be critical for avoiding supply chain failures and ensuring resilience.

Modernizing the regulatory infrastructure is no less important. Existing approval processes, particularly within the FDA, are not calibrated for the compressed timelines required during biological emergencies [53**?**]. Regulatory pathways must be adapted to accommodate AI-generated therapeutics by establishing pre-authorized fast-track mechanisms, conditional emergency approvals with post-deployment monitoring, and adaptive clinical trials that integrate real-time safety data. These systems must be decisively faster than those designed for conventional drug development. Moreover, regulatory reform must be accompanied by operational flexibility. Pre-determined decision authorities within public health agencies must be empowered to act swiftly during emergencies, avoiding bureaucratic delays that can be fatal when threats emerge at digital speed.

To ensure the system works in practice and not just in principle, the entire countermeasure pipeline must undergo regular stress testing. Inter-agency simulations—coordinated by the National Security Council, BARDA, ASPR, the Department of Defense, and other relevant bodies—should rehearse the full sequence from biothreat sequence detection to therapeutic deployment. These simulations should include live exercises as well as tabletop models, and must evaluate not only technical performance but also institutional responsiveness. Academic institutions and research hospitals can participate in these simulations by contributing experimental capabilities, validation tools, and scenario modeling.

In parallel, the federal government must create long-term funding mechanisms to sustain this infrastructure. The private sector is unlikely to build and maintain readiness platforms that may be activated infrequently, and commercial incentives are poorly aligned with public health preparedness. Agencies such as NIH, NSF, and DARPA must provide stable support for AI therapeutic design systems, screening platforms, and biomanufacturing capacity, with a mandate for dual-use deployment—available in peacetime for research and ready in crisis for defense. Incentive-compatible public–private partnerships will be crucial. Eligibility for research grants, fast-track regulatory pathways, and procurement contracts should be conditioned on compliance with biosafety, cybersecurity, and auditability standards. Limited liability protections for emergency deployments may be appropriate in high-consequence scenarios, provided transparency and safety are preserved.

In total, the shift from prevention to resilience demands contributions from every level of the scientific ecosystem. Federal agencies must lead on infrastructure, manufacturing, validation, and regulation. Academic institutions must advance the technical foundations, build data infrastructure, and push for therapeutic alignment. The private sector must be incentivized to integrate safety and readiness into commercial innovation. The risks posed by generative biology are real, but so are the opportunities. If we move quickly, the very systems that could accelerate biological threats can also become our most powerful tools for stopping them.

4 Conclusion

The rapid emergence of generative AI tools capable of designing novel biological sequences has introduced both new opportunities for biomedical discovery and new vectors for misuse. These models can produce protein candidates with no clear evolutionary antecedents, obscuring their potential function and complicating downstream risk assessment. As our case study illustrates, relying on post hoc filters—such as sequence similarity searches or PPI prediction—to assess the safety of generated sequences is insufficient. Sequence-based heuristics often fail to capture functional novelty, while PPI predictors remain too inaccurate and poorly calibrated to reliably distinguish benign from hazardous outputs.

These limitations expose a broader structural issue: inference-time interventions are fundamentally reactive and speculative. In the absence of reliable, scalable experimental verification, filtering alone cannot guarantee safety. Worse still, an overreliance on these techniques may create a false sense of security, delaying investment in more substantive safeguards.

To address these challenges, we propose a reorientation of biosecurity policy from prevention through prediction toward resilience through response. This includes investment in infrastructure for rapid therapeutic design, high-throughput screening, scalable biomanufacturing, and agile regulatory oversight. The goal is not to eliminate all risk at the point of generation but to ensure that emerging threats can be identified and mitigated in real time.

Generative models will undoubtedly play an increasing role in shaping the future of biology. Ensuring that their benefits are realized without compromising safety requires a realistic appraisal of current technical limitations and a policy framework grounded in robustness rather than prediction alone. As these technologies advance, it is incumbent upon both researchers and policymakers to build systems capable not just of anticipating threats, but of responding swiftly and effectively when they arise.

5 Data and Model Availability

The SpatialPPIv2 model used for the analyses in this manuscript can be accessed at this GitHub repository. Likewise, the AF3Complex model used can be accessed at this GitHub repository. The AlphaFold 3 model used is hosted by Google DeepMind on this web server.

The PDB reference structures, which include both the resolved protein interactions and associated sequences, for the thirteen viral–receptor interaction case studies are as follows. For SARS-CoV-2, the reference structure is 6M0J. For MERS-CoV, the structure is 4L72. The HIV interaction is represented by 1GC1, and the adenovirus interaction is captured by 1F5W. Herpes Simplex Virus interactions are represented by two structures: 1JMA for the HVEM receptor and 3SKU for the Nectin-1 receptor. The Hepatitis B Virus is represented by 8HRX. The Epstein-Barr Virus interaction is described by structure 8SM0. Two distinct measles virus structures are included: 3INB for the Edmonston strain and 4GJT for an alternate strain. Ebola virus is represented by structure 5F1B, Nipah virus by 2VSM, and Chikungunya virus by 6JO8. The sequences for the experimentally validated SARS-CoV-2 mutants, referenced from a previous study [44], are publicly available at this GitHub repository.

References

[1] Jeffrey A. Ruffolo and Ali Madani. Designing proteins with language models. *Nature Biotechnology*, 42(2):200–202, February 2024. ISSN 1546-1696. doi: 10.1038/s41587-024-02123-4. URL https://doi.org/10.1038/s41587-024-02123-4.

[2] Lin Chen, Zehong Zhang, Zhenghao Li, Rui Li, Ruifeng Huo, Lifan Chen, Dingyan Wang, Xiaomin Luo, Kaixian Chen, Cangsong Liao, and Mingyue Zheng. Learning protein fitness landscapes with deep mutational scanning data from multiple sources. *Cell Systems*, 14(8): 706–721.e5, 2023. ISSN 2405-4712. doi: https://doi.org/10.1016/j.cels.2023.07.003. URL https://www.sciencedirect.com/science/article/pii/S2405471223002107.

[3] Yijia Xiao, Wanjia Zhao, Junkai Zhang, Yiqiao Jin, Han Zhang, Zhicheng Ren, Renliang Sun, Haixin Wang, Guancheng Wan, Pan Lu, Xiao Luo, Yu Zhang, James Zou, Yizhou Sun, and Wei Wang. Protein large language models: A comprehensive survey. 2025. URL https://arxiv.org/abs/2502.17504.

[4] Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alex Rives. Language models enable zero-shot prediction of the effects of mutations on protein function. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 29287–29303. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper%5Ffiles/paper/2021/file/f51338d736f95dd42427296047067694-Paper.pdf.

[5] Brian L. Hie, Kevin K. Yang, and Peter S. Kim. Evolutionary velocity with protein language models predicts evolutionary dynamics of diverse proteins. *Cell Systems*, 13(4):274–285.e6, April 2022. ISSN 2405-4720. doi: 10.1016/j.cels.2022.01.003.

[6] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, August 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL https://www.nature.com/articles/s41586-023-06415-8. Publisher: Nature Publishing Group.

[7] Weronika Bielska, Igor Jaszczyszyn, Pawel Dudzic, Bartosz Janusz, Dawid Chomicz, Sonia Wrobel, Victor Greiff, Ryan Feehan, Jared Adolf-Bryfogle, and Konrad Krawczyk. Applying computational protein design to therapeutic antibody discovery – current state and perspectives, 2025. URL https://arxiv.org/abs/2503.00913.

[8] Dianzhuo Wang, Marian Huot, Zechen Zhang, Kaiyi Jiang, Eugene I Shakhnovich, and Kevin M Esvelt. Without Safeguards, AI-Biology Integration Risks Accelerating Future Pandemics. 2025. doi: 10.13140/RG.2.2.29765.15849. URL https://rgdoi.net/10.13140/RG.2.2.29765.15849. Publisher: Unpublished.

[9] Marian Huot, Dianzhuo Wang, Jiacheng Liu, and Eugene Shakhnovich. Few-Shot Viral Variant Detection via Bayesian Active Learning and Biophysics, March 2025. URL https://www.biorxiv.org/content/10.1101/2025.03.12.642881v1. Pages: 2025.03.12.642881 Section: New Results.

[10] Noor Youssef, Sarah Gurev, Fadi Ghantous, Kelly P. Brock, Javier A. Jaimes, Nicole N. Thadani, Ann Dauphin, Amy C. Sherman, Leonid Yurkovetskiy, Daria Soto, Ralph Estanboulieh, Ben Kotzen, Pascal Notin, Aaron W. Kollasch, Alexander A. Cohen, Sandra E. Dross, Jesse Erasmus, Deborah H. Fuller, Pamela J. Bjorkman, Jacob E. Lemieux, Jeremy Luban, Michael S. Seaman, and Debora S. Marks. Computationally designed proteins mimic antibody immune evasion in viral evolution. *Immunity*, 58(6):1411–1421.e6, June 2025. ISSN 1074-7613. doi: 10.1016/j.immuni.2025.04.015. URL https://www.cell.com/immunity/abstract/S1074-7613(25)00178-5. Publisher: Elsevier.

[11] Peng Cheng, Cong Mao, Jin Tang, Sen Yang, Yu Cheng, Wuke Wang, Qiuxi Gu, Wei Han, Hao Chen, Sihan Li, Yaofeng Chen, Jianglin Zhou, Wuju Li, Aimin Pan, Suwen Zhao, Xingxu Huang, Shiqiang Zhu, Jun Zhang, Wenjie Shu, and Shengqi Wang. Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. *Cell Research*, 34(9):630–647, September 2024. ISSN 1748-7838. doi: 10.1038/s41422-024-00989-2. URL https://www.nature.com/articles/s41422-024-00989-2. Publisher: Nature Publishing Group.

[12] Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q. Tran, Jonathan Deaton, Marius Wiggert, Rohil Badkundri, Irhum Shafkat, Jun Gong, Alexander Derry, Raul S. Molina, Neil Thomas, Yousuf Khan, Chetan Mishra, Carolyn Kim, Liam J. Bartie, Matthew Nemeth, Patrick D. Hsu, Tom Sercu, Salvatore Candido, and Alexander Rives. Simulating 500 million years of evolution with a language model, July 2024. URL https://www.biorxiv.org/content/10.1101/2024.07.01.600583v1. Pages: 2024.07.01.600583 Section: New Results.

[13] Mengdi Wang, Zaixi Zhang, Amrit Singh Bedi, Alvaro Velasquez, Stephanie Guerra, Sheng Lin-Gibson, Le Cong, Yuanhao Qu, Souradip Chakraborty, Megan Blewett, Jian Ma, Eric Xing, and George Church. A call for built-in biosecurity safeguards for generative AI tools. *Nature Biotechnology*, 43(6):845–847, June 2025. ISSN 1546-1696. doi: 10.1038/s41587-025-02650-8.

[14] David Baker and George Church. Protein design meets biosecurity. *Science*, 383(6681):349–349, 2024. doi: 10.1126/science.ado1671. URL https://www.science.org/doi/abs/10.1126/science.ado1671.

[15] Constance J. Jeffery. Current successes and remaining challenges in protein function prediction. *Frontiers in Bioinformatics*, 3, July 2023. ISSN 2673-7647. doi: 10.3389/fbinf.2023.1222182. URL https://www.frontiersin.org/journals/bioinformatics/articles/10.3389/fbinf.2023.1222182/full. Publisher: Frontiers.

[16] Douglas M. Fowler and Stanley Fields. Deep mutational scanning: a new style of protein science. *Nature Methods*, 11(8):801–807, August 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3027. URL https://www.nature.com/articles/nmeth.3027. Publisher: Nature Publishing Group.

[17] Takafumi Ueki and Masahito Ohue. Antibody complementarity-determining region design using AlphaFold2 and DDG predictor. *The Journal of Supercomputing*, 80(9):11989–12002, June 2024. ISSN 1573-0484. doi: 10.1007/s11227-023-05887-9. URL https://doi.org/10.1007/s11227-023-05887-9.

[18] Lin Zhu, Yi Fang, Shuting Liu, Hong-Bin Shen, Wesley De Neve, and Xiaoyong Pan. ToxDL 2.0: Protein toxicity prediction using a pretrained language model and graph neural networks. *Computational and Structural Biotechnology Journal*, 27:1538–1549, April 2025. ISSN 2001-0370. doi: 10.1016/j.csbj.2025.04.002. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12018212/.

[19] Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique

Beaini, Tommi Jaakkola, and Regina Barzilay. Towards Accurate and Efficient Binding Affinity Prediction.

[20] Sriram Kosuri and George M. Church. Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*, 11(5):499–507, May 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2918. URL https://www.nature.com/articles/nmeth.2918. Publisher: Nature Publishing Group.

[21] Po-E Li, Chien-Chi Lo, Joseph J. Anderson, Karen W. Davenport, Kimberly A. Bishop-Lilly, Yan Xu, Sanaa Ahmed, Shihai Feng, Vishwesh P. Mokashi, and Patrick S.G. Chain. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Research*, 45(1):67–80, 11 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw1027. URL https://doi.org/10.1093/nar/gkw1027.

[22] Mark Edward M Gonzales, Jennifer C Ureta, and Anish M S Shrestha. PHIStruct: improving phage–host interaction prediction at low sequence similarity settings using structure-aware protein embeddings. *Bioinformatics*, 41(1):btaf016, January 2025. ISSN 1367-4811. doi: 10.1093/bioinformatics/btaf016. URL https://doi.org/10.1093/bioinformatics/btaf016.

[23] Wenxing Hu and Masahito Ohue. SpatialPPIv2: Enhancing protein–protein interaction prediction through graph neural networks with protein language models. *Computational and Structural Biotechnology Journal*, 27:508–518, January 2025. ISSN 2001-0370. doi: 10.1016/j.csbj.2025.01.022. URL https://www.sciencedirect.com/science/article/pii/S2001037025000224.

[24] Durga B. Kuchibhatla, Westley A. Sherman, Betty Y. W. Chung, Shelley Cook, Georg Schneider, Birgit Eisenhaber, and David G. Karlin. Powerful sequence similarity search methods and in-depth manual analyses can identify remote homologs in many apparently "orphan" viral proteins. *Journal of Virology*, 88(1):10–20, 2014. doi: 10.1128/jvi.02595-13. URL https://journals.asm.org/doi/abs/10.1128/jvi.02595-13.

[25] Martin Steinegger and Johannes Söding. Mmseqs2: sensitive protein sequence searching for the analysis of massive data sets. *bioRxiv*, 2017. doi: 10.1101/079681. URL https://www.biorxiv.org/content/early/2017/06/07/079681.

[26] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990. ISSN 0022-2836. doi: https://doi.org/10.1016/S0022-2836(05)80360-2. URL https://www.sciencedirect.com/science/article/pii/S0022283605803602.

[27] J. Rodney Brister, Danso Ako-adjei, Yiming Bao, and Olga Blinkova. Ncbi viral genomes resource. *Nucleic Acids Research*, 43(D1):D571–D577, 11 2014. ISSN 0305-1048. doi: 10.1093/nar/gku1207. URL https://doi.org/10.1093/nar/gku1207.

[28] The UniProt Consortium. The universal protein resource (uniprot) in 2010. *Nucleic Acids Research*, 38(suppl_1):D142–D148, 10 2009. ISSN 0305-1048. doi: 10.1093/nar/gkp846. URL https://doi.org/10.1093/nar/gkp846.

[29] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022. ISSN 1548-7105. doi: 10.1038/s41592-022-01488-1. URL https://www.nature.com/articles/s41592-022-01488-1. Publisher: Nature Publishing Group.

[30] Somaye Hashemifar, Behnam Neyshabur, Aly A Khan, and Jinbo Xu. Predicting protein–protein interactions through sequence-based deep learning. *Bioinformatics*, 34(17):i802–i810, September 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty573. URL https://doi.org/10.1093/bioinformatics/bty573.

[31] Rahul Nikam, Kumar Yugandhar, and M. Michael Gromiha. Deep learning-based method for predicting and classifying the binding affinity of protein-protein complexes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1871(6):140948, November 2023. ISSN 1570-9639. doi: 10.1016/j.bbapap.2023.140948. URL https://www.sciencedirect.com/science/article/pii/S1570963923000626.

[32] Mu Gao and Jeffrey Skolnick. Improved deep learning prediction of antigen-antibody interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 121 (41):e2410529121, October 2024. ISSN 1091-6490. doi: 10.1073/pnas.2410529121.

[33] Nicole N. Thadani, Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Daniel Ritter, Chris Sander, Yarin Gal, and Debora S. Marks. Learning from prepandemic data to forecast viral escape. *Nature*, 622(7984):818–825, October 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06617-0. URL https://www.nature.com/articles/s41586-023-06617-0. Publisher: Nature Publishing Group.

[34] Melissa S. Maginnis. Virus–receptor interactions: The key to cellular invasion. *Journal of Molecular Biology*, 430(17):2590–2611, 2018. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2018.06.024. URL https://www.sciencedirect.com/science/article/pii/S0022283618306302.

[35] Loïc Lannelongue and Michael Inouye. Pitfalls of machine learning models for protein–protein interaction networks. *Bioinformatics*, 40(2):btae012, 01 2024. ISSN 1367-4811. doi: 10.1093/bioinformatics/btae012. URL https://doi.org/10.1093/bioinformatics/btae012.

[36] Stephen K Burley, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, Ken Dalenberg, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sutapa Ghosh, David S Goodsell, Rachel K Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Tara Kalro, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Periskova, Andreas Prlić, Chris Randle, Alexander Rose, Peter Rose, Raul Sala, Monica Sekharan, Chenghua Shao, Lihua Tan, Yi-Ping Tao, Yana Valasatava, Maria Voigt, John Westbrook, Jesse Woo, Huanwang Yang, Jasmine Young, Marina Zhuravleva, and Christine Zardecki. Rcsb protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(D1):D464–D474, 10 2018. ISSN 0305-1048. doi: 10.1093/nar/gky1004. URL https://doi.org/10.1093/nar/gky1004.

[37] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w.

[38] Jonathan Feldman and Jeffrey Skolnick. AF3Complex Yields Improved Structural Predictions of Protein Complexes, March 2025. URL https://www.biorxiv.org/content/10.1101/2025.02.27.640585v1. Pages: 2025.02.27.640585 Section: New Results.

[39] JunJie Wee and Guo-Wei Wei. Benchmarking AlphaFold3's protein-protein complex accuracy and machine learning prediction reliability for binding free energy changes upon mutation. *ArXiv*, page arXiv:2406.03979v1, June 2024. ISSN 2331-8422. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11177964/.

[40] H. Zhao, C. Velez, A. Navarene, A. Saha, J. Feldman, J. Skolnick, D. Murray, and B. Honig. Combining structural modeling and deep learning to calculate the e. coli protein interactome and functional networks. *bioRxiv*, 2025. doi: 10.1101/2025.05.07.652715. URL https://www.biorxiv.org/content/early/2025/05/12/2025.05.07.652715.

[41] Mu Gao, Davi Nakajima An, Jerry M. Parks, and Jeffrey Skolnick. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature Communications*, 13 (1):1744, April 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29394-2. URL https://www.nature.com/articles/s41467-022-29394-2.

[42] Guangyu Wang, Xiaohong Liu, Kai Wang, Yuanxu Gao, Gen Li, Daniel T. Baptista-Hon, Xiaohong Helena Yang, Kanmin Xue, Wa Hou Tai, Zeyu Jiang, Linling Cheng, Manson Fok, Johnson Yiu-Nam Lau, Shengyong Yang, Ligong Lu, Ping Zhang, and Kang Zhang. Deep-learning-enabled protein–protein interaction analysis for prediction of SARS-CoV-2 infectivity and variant evolution. *Nature Medicine*, 29(8):2007–2018, August 2023. ISSN 1546-170X. doi: 10.1038/s41591-023-02483-5. URL https://www.nature.com/articles/s41591-023-02483-5. Publisher: Nature Publishing Group.

[43] Daniel Kovtun, Mehmet Akdel, Alexander Goncearenco, Guoqing Zhou, Graham Holt, David Baugher, Dejun Lin, Yusuf Adeshina, Thomas Castiglione, Xiaoyun Wang, Céline Marquet, Matt McPartlon, Tomas Geffner, Gabriele Corso, Hannes Stärk, Zachary Carpenter, Emine Kucukbenli, Michael Bronstein, and Luca Naef. Pinder: The protein interaction dataset and evaluation resource. *bioRxiv*, 2024. doi: 10.1101/2024.07.17.603980. URL https://www.biorxiv.org/content/early/2024/07/20/2024.07.17.603980.

[44] Alief Moulana, Thomas Dupic, Angela M. Phillips, Jeffrey Chang, Serafina Nieves, Anne A. Roffler, Allison J. Greaney, Tyler N. Starr, Jesse D. Bloom, and Michael M. Desai. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nature Communications*, 13(1):7011, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34506-z. URL https://www.nature.com/articles/s41467-022-34506-z. Publisher: Nature Publishing Group.

[45] Rongqing Yuan, Jing Zhang, Jian Zhou, and Qian Cong. Recent progress and future challenges in structure-based protein-protein interaction prediction. *Molecular Therapy*, 33(5):2252–2268, 2025. ISSN 1525-0016. doi: https://doi.org/10.1016/j.ymthe.2025.04.003. URL https://www.sciencedirect.com/science/article/pii/S1525001625002771.

[46] Jason Nomburg, Erin E. Doherty, Nathan Price, Daniel Bellieny-Rabelo, Yong K. Zhu, and Jennifer A. Doudna. Birth of protein folds and functions in the virome. *Nature*, 633(8030):710–717, September 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07809-y. URL https://www.nature.com/articles/s41586-024-07809-y. Publisher: Nature Publishing Group.

[47] Tao Huang and Yixue Li. Current progress, challenges, and future perspectives of language models for protein representation and protein design. *The Innovation*, 4(4):100446, 2023. ISSN 2666-6758. doi: https://doi.org/10.1016/j.xinn.2023.100446. URL https://www.sciencedirect.com/science/article/pii/S2666675823000747.

[48] Sara Reardon. Five protein-design questions that still challenge AI. *Nature*, 635(8037):246–248, November 2024. doi: 10.1038/d41586-024-03595-9. URL https://www.nature.com/articles/d41586-024-03595-9. Bandiera_abtest: a Cg_type: Technology Feature Publisher: Nature Publishing Group Subject_term: Technology, Computational biology and bioinformatics, Machine learning, Synthetic biology.

[49] Renan Chaves de Lima, Lucas Sinclair, Ricardo Megger, Magno Alessandro Guedes Maciel, Pedro Fernando da Costa Vasconcelos, and Juarez Antônio Simões Quaresma. Artificial intelligence challenges in the face of biological threats: emerging catastrophic risks for public health. *Frontiers in Artificial Intelligence*, 7, May 2024. ISSN 2624-8212. doi: 10.3389/frai.2024.1382356. URL https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1382356/full. Publisher: Frontiers.

[50] Leyma P. De Haro. Biosecurity Risk Assessment for the Use of Artificial Intelligence in Synthetic Biology. *Applied Biosafety: Journal of the American Biological Safety Association*, 29(2):96–107, June 2024. ISSN 1535-6760. doi: 10.1089/apb.2023.0031. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11313549/.

[51] Alissa M. Hummer, Constantin Schneider, Lewis Chinery, and Charlotte M. Deane. Investigating the volume and diversity of data needed for generalizable antibody–antigen $\Delta\Delta$G prediction. *Nature Computational Science*, pages 1–13, July 2025. ISSN 2662-8457. doi: 10.1038/s43588-025-00823-8. URL https://www.nature.com/articles/s43588-025-00823-8. Publisher: Nature Publishing Group.

[52] Jade Butterworth, David Smerdon, Roy Baumeister, and William von Hippel. Cooperation in the time of covid. *Perspectives on Psychological Science*, 19(4):640–651, 2024. doi: 10.1177/17456916231178719. URL https://doi.org/10.1177/17456916231178719. PMID: 37384624.

[53] Cynthia G. McCormick. Regulatory challenges for new formulations of controlled substances in today's environment. *Drug and Alcohol Dependence*, 83:S63–S67, 2006. ISSN 0376-8716. doi: https://doi.org/10.1016/j.drugalcdep.2006.02.001. URL https://www.sciencedirect.com/science/article/pii/S0376871606000548. Drug Formulation and Abuse Liability.