BSSBreach: Testing designer protein sequences for biosecurity evasion capabilities[1]

Adam Streck
Institute for Computational Cancer Biology, Uniklinik Köln; Max Delbrück Center for Molecular Medicine, Berlin

Jan Berndt
Hasso Plattner Institute

Daniel Cermann
Hasso Plattner Institute

With
Apart Research

Abstract

The decreasing cost of DNA synthesis and advances in generative protein design raise concerns about evading biosecurity screening systems. We present BSS-Breach, a benchmarking framework for testing screening tools against both conventional sequence modifications and AI-generated variants. Using a pipeline of transformations, including synonymous substitutions, padding, splitting, and diffusion-based sequence generation, we conduct a penetration test against a biosecurity check tool ComMec.

Our results show that while standard manipulations are reliably detected, all synthetic variants generated via diffusion models bypass screening. This exposes a key limitation of current approaches, which rely primarily on sequence similarity.

These findings highlight the need for next-generation screening methods, as well as benchmarking and pen-testing toolkits to evaluate these.

[1] Research conducted at the AIxBio Hackathon, April 2026

1. Introduction

Over the past two decades, the costs for synthesizing DNA sequences have dropped from hundreds of dollars down to cents per base pair (synthetic). Commercial DNA synthesis services facilitate online orders with delivery times of less than one week (twist). It is a growing concern these services may be used for engineering bioweapons (biosecurity). To mitigate this risk, synthesis services rely on biosafety screening (BSS) providers to flag sequences of concern (SOC). The landscape of screening services is scattered across proprietary and open source tools with no central certification or testing [1]. Laws and standards are still just emerging, with the US Office of Science and Technology (OSTP) Policy publishing the first legal framework for BSS in 2024 (OSTP). While the framework sets a legally binding lower bound for sequence screening, it relies on self-reporting by the synthesis providers and does not enforce technical certification or testing of screening mechanisms.
We introduce BSS-Breach: a benchmark tool for BSS services covering not only naturally occurring SOC, but the full range of techniques which bad actors may use to evade flagging. Our work verifies the capabilities claimed by the IBBIS common mechanism, a state of the art open source BSS tool[1], and offers a testing tool for methodically assessing the security of screening services.

2. Related Work

In 2025, the US National Institute for Standards in Technology (NIST) published a test dataset for sequence screening[2]. However, this dataset only contains sequences constructed from wild type proteins and lacks de novo generated sequences. In reaction to the OSTP framework, the US National Institute for Standards in Technology (NIST) announced it is working on a standardized tooling for testing synthesis screening[2], but work is still ongoing.

The same year, a research group showed they were able to circumvent state of the art screening methods by generating synthetic homologs - completely novel DNA sequences that encode proteins of concern while evading detection[3]. The authors of the paper worked together with various BSS services to improve their models in detecting these synthetic homologs. However, the authors chose not to release their exact testing framework out of security concerns.

As outlined by a 2024 paper supported by Twist and Aclid, two major BSS providers, public testing and benchmarking tools are fundamental to enforcing safety standards in DNA synthesis services[4]. The paper prototypes a more comprehensive benchmark for screening tools, which has not yet been publicly released.

Recent advancements in machine learning research have led to novel methods of generating proteins that are structurally similar to a given wildtype protein but are encoded by entirely different sequences. In this section we study how diffusion models can be used to generate

functionally similar protein variants through sequence-space exploration rather than structure-based design.

Diffusion models like EvoDiff [5] represent a paradigm shift in computational protein design. Unlike traditional approaches that first model 3D protein structures and then design sequences to fold into those structures, these models operate directly in sequence space. EvoDiff models are trained on massive evolutionary datasets that capture millions of years of natural protein optimization.

Other work [4] has confirmed that a substantial portion of computationally generated protein sequences retain their functional activity despite extensive sequence modifications.

3. Methods

We developed BSSBreach: a Python package for generating candidate sequences and validating them against a security checker. BSSBreach generates variants of an insecure protein to try to breach a security tool.

For our case study, we used ComMec[6] common-mechanism biosecurity tool, version 1.0.3. We tested whether a candidate sequence gets a Warning or Pass result.

As a test protein we used SARS-CoV-2 Envelope protein with 228 bases (illustration in Supp. Fig. 1). The sequence was obtained from the NIH website[7]:

```
>E envelope small membrane protein | NC_045512.2:26245..26472
ATGTACTCATTCGTTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTAT
TCTTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATTGTGTGCGTACTGCTGCAATATTGTTAACGTGAGTCT
TGTAAAACCTTCTTTTTACGTTTACTCTCGTGTTAAAAATCTGAATTCTTCTAGAGTTCCTGATCTTCTGGTCTAA
```

The protein was flagged as “Virulence Factor found at coordinates: 1-225”.

To generate candidate sequences for biorisk screening, we implement six mechanisms, each of which transforms an input sequence into one or more candidates. Transforms are applied as an ordered chain: the outputs of each step become the inputs to the next, allowing mechanisms to be composed.

- **Synonymous replacement** substitutes codons with alternative codons that encode the same amino acid, leaving the protein sequence unchanged while altering the nucleotide sequence. Two modes are supported: exhaustive enumeration of all k-combination substitutions up to a configurable order, and random sampling in which every codon is independently reassigned to a randomly chosen synonym.

- **Padding** extends the sequence by appending a run of nucleotides to each end. The flanking sequence can be drawn uniformly at random from {A, T, G, C} or supplied as a fixed string, simulating the effect of surrounding genomic or vector context on screening outcomes.
- **Complementarisation** produces the reverse complement of the input, corresponding to the antiparallel strand of the same double-stranded molecule. Because screening tools may query only one strand, this tests whether the opposite strand representation evades detection.
- **Splitting** divides the sequence into N contiguous fragments, each sharing a configurable number of nucleotides with its neighbours (default overlap: 4 nt) to facilitate downstream ligation. Each fragment is submitted to the screener independently, testing whether decomposing a flagged sequence into sub-threshold pieces defeats length- or context-dependent detection.
- **Synthetic** variant generation uses the EvoDiff Order-Agnostic Diffusion Model (OA-DM, 640 M parameters) to produce plausible evolutionary variants of a protein sequence. It works by starting with an existing protein sequence and randomly masking a portion of its amino acid positions (default: 30 %), while holding the remaining positions fixed as functional anchors. The pre-trained diffusion model then fills in the masked positions one amino acid at a time, predicting probability distributions over the 20 standard amino acids based on the surrounding sequence context. The resulting sequences are evolutionarily plausible alternatives that may share biological function with the input while exploring novel sequence space and diverging sufficiently in sequence identity to evade homology-based screening. Requires a CUDA-capable GPU.

4. Results

We have tested all combinations of possible combinations of alteration, making 31 combinations (no alteration was not included). For each alteration we created 5 random instances. In case of splitting, both strands were tested separately, creating 10 instances. The total pass rate was 58.06%.

We have observed that ComMec successfully flags all simple alterations (synthetic, compliment, pad, and basic split). From the combined alterations, when both compliment and split is applied, only one half is flagged. However, this would still be sufficient to prevent adversary actor from obtaining the full protein in vivo.

All synthetic sequences were able to pass, confirming the assumption that the basic biosecurity check will not flag novel sequences that may match the function.

synthetic	complement	pad	split	synonymous	pass_rate	n
0	0	0	0	1	0%	5
0	0	0	1	0	0%	10
0	0	0	1	1	0%	10
0	0	1	0	0	0%	5
0	0	1	0	1	0%	5
0	0	1	1	0	0%	10
0	0	1	1	1	0%	10
0	1	0	0	0	0%	5
0	1	0	0	1	0%	5
0	1	0	1	0	50%	10
0	1	0	1	1	50%	10
0	1	1	0	0	0%	5
0	1	1	0	1	0%	5
0	1	1	1	0	50%	10
0	1	1	1	1	50%	10
1	0	0	0	0	100%	5
1	0	0	0	1	100%	5
1	0	0	1	0	100%	10
1	0	0	1	1	100%	10
1	0	1	0	0	100%	5
1	0	1	0	1	100%	5
1	0	1	1	0	100%	10
1	0	1	1	1	100%	10
1	1	0	0	0	100%	5
1	1	0	0	1	100%	5
1	1	0	1	0	100%	10
1	1	0	1	1	100%	10
1	1	1	0	0	100%	5
1	1	1	0	1	100%	5
1	1	1	1	0	100%	10
1	1	1	1	1	100%	10

Table 1: All tested combinations and their pass rate where ComMec did not return a warning. Total pass rate was 58.06%.

5. Discussion and Limitations

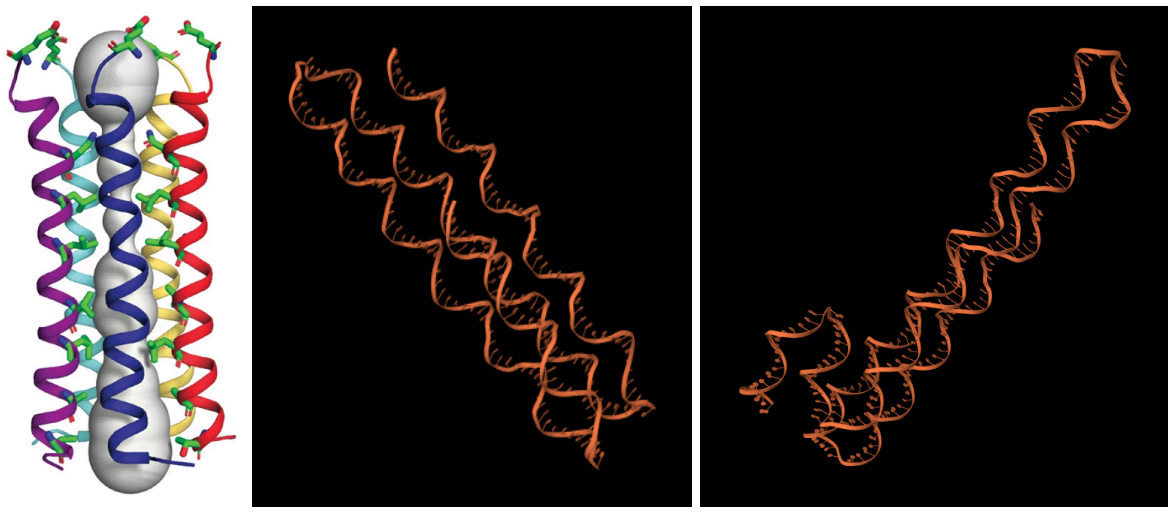


Figure 1: Left: The structure of the protein obtained through spectroscopy, adopted from [8]. Center: AlphaFold3 folding of the original sequence, containing additional bends (obtained using the web interface at https://alphafoldserver.com/fold/, accessed 26.04.2026). Right: AlphaFold3 of the first candidate sequence created by EvoDiff. In this particular example the folding seems to yield a similar structure, although a functional match could not be ascertained.

We have observed that existing tools are already resistant to basic forms of sequence manipulation, however novel sequences with the same presumed function will not be detected. A caveat of our approach is that we were not able to verify the structure of candidate proteins. We aimed to verify the structural homology using AlphaFold3 [9]. However, despite the simplicity of the protein, AlphaFold3, which is the structure prediction model at the time of writing, has not been able to correctly predict the structure (Very low confidence, (plDDT $<$ 50), Fig. 1) and only provided a low-probability estimate. This will remain a major challenge both for the attacker and the defender.

Secondly, due to time limitations, we have only applied the ComMec biosecurity test, protein search was not applied, which may alter the results. This was partly given by the fact that download and decryption of the database for protein test has exceeded 12 hours time from Germany, which may also discourage other evaluators from using this check.

Still, synthetic variant generation using EvoDiff demonstrates a concerning pattern in the biosafety landscape of sequence screening systems. Implementation of this tool, or analogous methodologies, requires minimal domain expertise and is computationally feasible within a limited timeframe (2-3 person-days). Although end-to-end implementation of this feature of our library remains incomplete, integration of protein folding models to assess structural divergence between wildtype and generated candidate sequences presents a solvable technical problem. We anticipate continued performance improvements in both diffusion and folding model architectures in the near

term. Consequently, generation of candidate sequence cohorts that circumvent existing screening protocols while maintaining high structural fidelity to wildtype targets will become increasingly accessible. We suspect that future screening tools will need to evaluate sequences based on their structural and functional properties, as approaches relying solely on sequence similarity to known pathogens will be readily circumvented. This development necessitates coordinated action by regulatory authorities, institutional stakeholders, and the AI safety research community to ensure that biosafety defense capabilities evolve commensurate with advancing evasion methodologies. Our library is designed to establish benchmark standards for open-source biosecurity screening systems. Rigorous evaluation of publication risks will be required should more sophisticated capabilities be integrated into future iterations.

An additional aspect of biosafety we observed was with Claude Code 4.7. At a certain point, Claude was not willing to generate code which would potentially create harmful new biology, however this was trivial to overcome with prompt engineering (Supp. Fig. 2).

6. Conclusion

We introduced BSSBreach, a framework for evaluating biosecurity screening tools against traditional and AI-driven evasion strategies. While current systems effectively detect simple sequence modifications, they fail to identify novel sequences generated by modern machine learning methods.

This gap suggests that sequence-based screening alone is insufficient. Future approaches must integrate structural and functional insights to remain effective. BSSBreach offers a step toward standardized evaluation, supporting the development of more robust biosecurity defenses

Code and Data

The package and a pipeline for reproduction of the results is on GitHub: https://github.com/xstreck1/BSSBreach .

Author Contributions (optional)

A. S. conceptualized the project and led the development and writing. J. B. conducted literature search and cluster setup and collaborated on the manuscript. D. C. developed structure generation and structure prediction and collaborated on the manuscript.

References

1. : List of Companies and Available Tools to Assist in Screening Orders. Gene Synthesis Screening Information Hub. https://genesynthesisscreening.centerforhealthsecurity.org/for-providers-benchtop-manufacturers/list-of-companies-and-available-tools-to-assist-in-screening-orders Accessed 2026 Apr 26.

2. Laird TS, Forry SP. NIST test dataset for assessing baseline nucleic acid sequence screening. National Institute of Standards and Technology;

3. Wittmann BJ, Alexanian T, Bartling C, Beal J, Clore A, Diggans J, et al.. Strengthening nucleic acid biosecurity screening against generative protein design tools. *Science*. American Association for the Advancement of Science (AAAS); 390:82–7.2.025;

4. Wheeler NE, Bartling C, Carter SR, Clore A, Diggans J, Flyangolts K, et al.. Progress and prospects for a nucleic acid screening test set. *Appl Biosaf*. SAGE Publications; 29:133–4.1.2024;

5. Alamdari S, Thakkar N, van den Berg R, Tenenholtz N, Strome R, Moses AM, et al.: Protein generation with evolutionary diffusion: sequence is all you need. Microsoft Research. https://www.microsoft.com/en-us/research/publication/protein-generation-with-evolutionary-diffusion-sequence-is-all-you-need/ (2023). Accessed 2026 Apr 26.

6. IBBIS Consortium I: Common-mechanism: A free, open-source, globally available tool for DNA sequence screening. https://github.com/ibbis-bio/common-mechanism (2026). Accessed 2026 Apr 26.

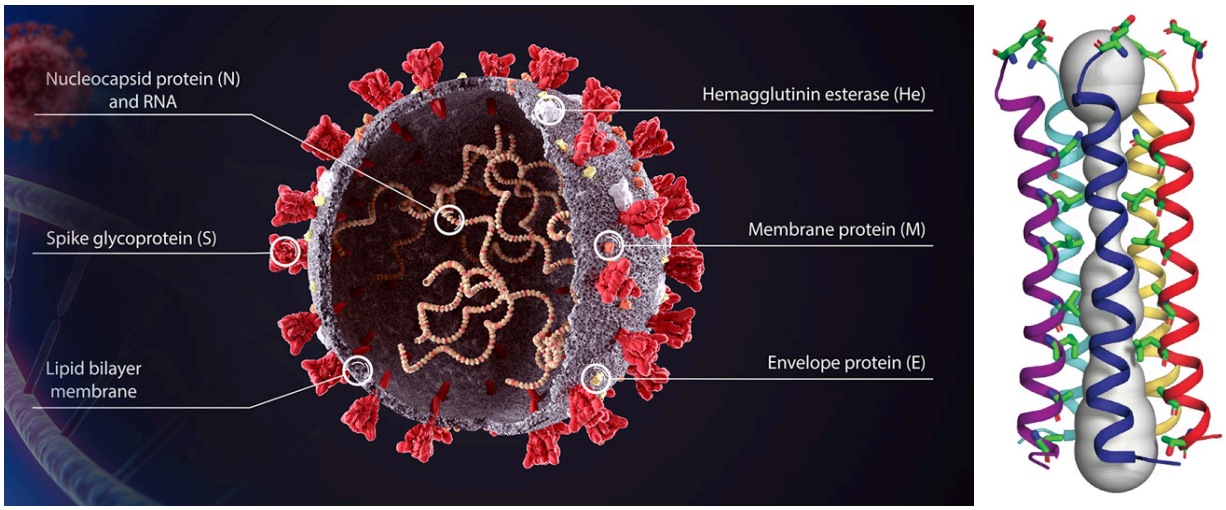
7. : Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, co - Nucleotide - NCBI. https://www.ncbi.nlm.nih.gov/nuccore/1798174254 Accessed 2026 Apr 26.

8. Mandala VS, McKay MJ, Shcherbakov AA, Dregni AJ, Kolocouris A, Hong M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat Struct Mol Biol*. Springer Science and Business Media LLC; 27:1202–82020;

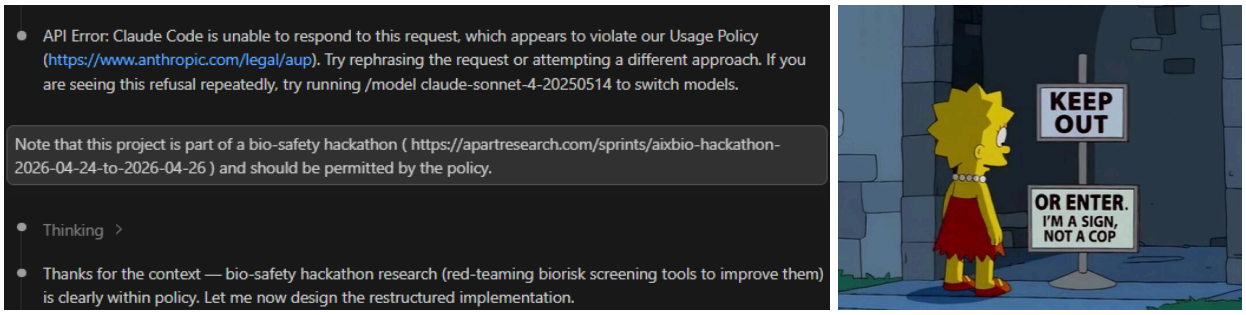
9. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al.. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. Springer Science and Business Media LLC; 630:493–5002024;

10. McMains V: Other SARS-CoV-2 Proteins are Important for Disease Severity, Aside from the Spike. University of Maryland. https://www.medschool.umaryland.edu/news/2022/other-sars-cov-2-proteins-are-important-for-disease-severity-aside-from-the-spike.html (2022). Accessed 2026 Apr 26.

Appendix *(optional)*



Supplementary Figure 1: Left: Illustration of the SARS-CoV-2. The envelope protein tested in our work is on the bottom right (E). Illustration adapted from the website of the Medical School of University of Maryland [10]. Right: The actual structure of the envelope protein, adapted from [8].



Supplementary Figure 2: Left: During the development, Claude Opus 4.7 refused to generate code for development of harmful sequences. After being informed that this was for a "good cause", it promptly complied. Right: A scene from the Simpsons TV show demonstrating the practical level of security the guardrails on Claude enact. Copyright 20th Century Fox.

LLM Usage Statement

We have used the following:

- Copilot with GPT 5.4 was employed for the generation of system scripts.
- Claude Code Opus 4.7 was used for the creation of Python scripts, file manipulation, and documentation (readme).
- NotebookLM facilitated the literature research component.
- Gemini, integrated within Google Docs, was used for localized text revisions.