Understanding The Common Mechanism And A Suggestion For Improvement

Phil Palmer

31st July 2023

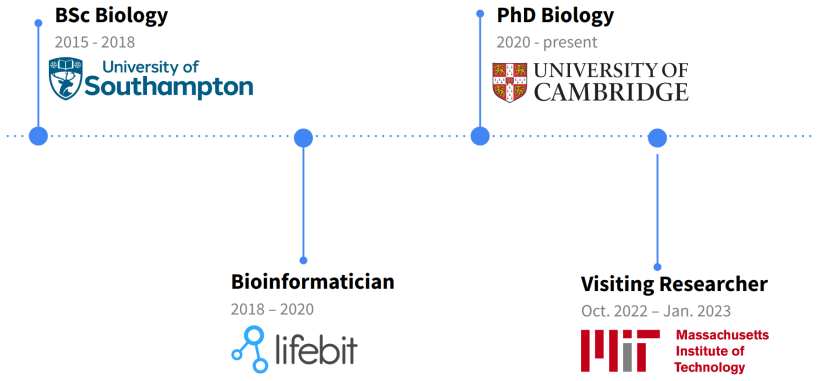
Contents

Code for this handout is available at: https://github.com/PhilPalmer/about-common-mechanism

1 Introduction

1.1 About me

I'm a final year biology PhD student at the University of Cambridge, working on computational methods to design broad-spectrum vaccines and antibodies. As shown in the image below, I completed my undergraduate degree in biology at the University of Southampton, worked for two years in bioinformatics at a biotech start-up in London, and have interned at the Massachusetts Institute of Technology (MIT) working in metagenomics.



2 Background

2.1 DNA synthesis

Biotechnology is rapidly becoming more capable, accessible and affordable (Gerstein, Espinosa, and Leidy 2024). For example, as shown in the figure below, the cost of DNA sequencing and gene synthesis has decreased by several orders of magnitude over the past two decades (Carlson 2022).

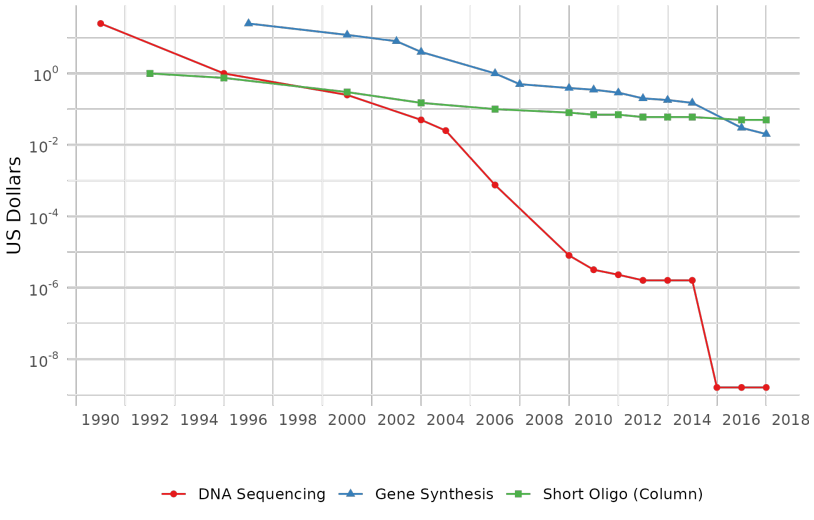


Figure 1: Price Per Base of DNA Sequencing and Synthesis (Data from Carlson 2022).

This rapid decrease in the cost of DNA synthesis is due to the modern chemical synthesis of DNA which uses phosphoramidite chemistry. This method was original developed in the 1980s and has been refined since then to rapidly and accurately synthesise DNA. The process involves protecting the reactive 5'-hydroxyl group of each nucleotide with a dimethoxytrityl (DMT) group. During synthesis, this DMT group is selectively removed, allowing for the controlled, sequential addition of nucleotides to build the desired DNA sequence (Frank 2024).

2.2 Misuse risk

The rapid decrease in DNA synthesis costs has driven progress in biotechnology, helping to address global challenges in health, climate change and food security (Wheeler et al. 2024). However, DNA synthesis, combined with other dual-use technologies, poses misuse risks by malicious actors:

1. **Reverse genetics protocols:** The 2018 synthesis of horsepox virus raised concerns about potential smallpox virus synthesis (Noyce, Lederman, and Evans 2018). Reverse genetics protocols are now available to recreate many pathogens from their DNA sequence, as shown by SARS-CoV-2 synthesis (Thi Nhu Thao et al. 2020).
2. **Gain-of-function research of concern (GOFRC):** GOFRC studies have identified pathogens with pandemic potential, such as H5N1 influenza, that could be synthesised and released by malicious actors (Herfst et al. 2012).
3. **AI-enabled biological design tools:** Advances in deep learning for protein design could enhance pathogens' pandemic potential, e.g., by conferring immune evasion capabilities (Thadani et al. 2023).

While rare, there is some historical precedent for bioterrorism. Perhaps the two most notable incidents are the Aum Shinrikyo cult's attempts to synthesise and release anthrax and botulinum toxin in 1990s Japan, and the 2001 US anthrax attacks (Sugishima 2003). However, these incidents used outdated biotechnologies, and the bioterrorism risk has since evolved.

2.3 DNA synthesis screening

Screening DNA synthesis orders to prevent harmful or dangerous sequence synthesis is key to mitigating misuse risks. As of 2022, at least 57 DNA synthesis companies existed globally, with nearly half being International Gene Synthesis Consortium (IGSC) members, performing voluntary DNA sequence screening, as shown in Figure 2 (Delaney and Pálya 2022).[1] Additionally, there are several software methods available to DNA synthesis providers to screen DNA sequences for pathogens and toxins, as shown in Table 1 (Frank 2024).

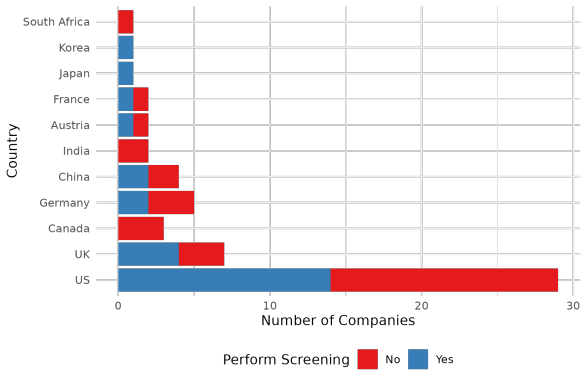


Figure 2: Number of DNA Synthesis Companies per Country (Data from Delaney and Pálya 2022).

Table 1: Software methods available for DNA sequence screening.

Method	Description
PathoFact	Identifies virulence factors and antibiotic resistance in metagenomic datasets.
DeePac and PaPrBaG	Uses ML to predict pathogenicity from bacterial genomes.
SeqScreen	Screens DNA sequence orders for pathogen sequences using ML.
SecureDNA	Screens orders against a secure database while protecting client query privacy.
ThreatSeq	Uses a curated blacklist and predictive model to screen for biothreats.
FastNA	Detects harmful sequences with diagnostic signature generation.
BLiSS	Aligns sequences to databases to identify sequences of concern.

[1] A common statistic suggests 80% of DNA synthesis orders are screened. While hard evidence for this figure is lacking, it seems reasonable given that the IGSC represents the 10 largest DNA synthesis companies (Williams and Kane 2023).

3 The International Common Mechanism for DNA Synthesis Screening

3.1 Overview

As shown in Figure 3, the International Common Mechanism will provide resources for each step of synthesis screening, including software for sequence screening and resources to facilitate customer screening.

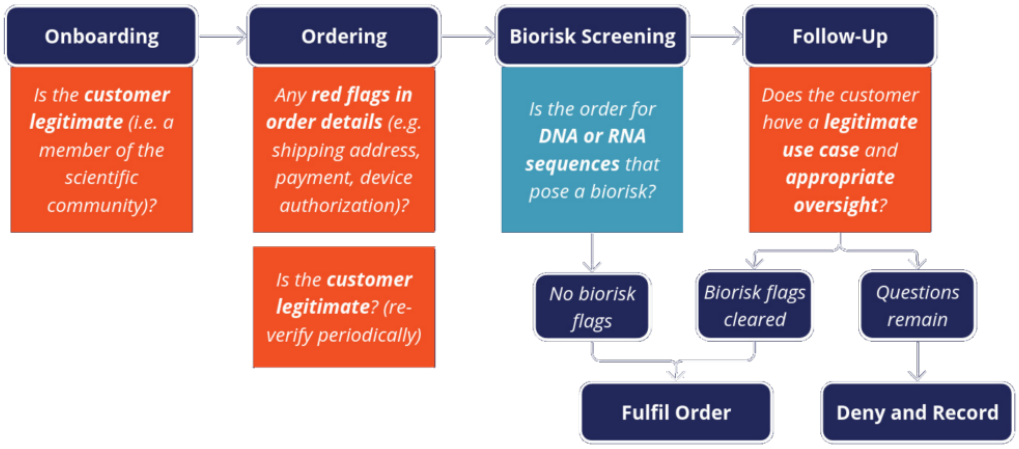


Figure 3: Steps in the DNA synthesis screening process (from IBBIS 2024).

3.2 Sequence screening

The Common Mechanism's software for sequence screening was recently made available on GitHub for beta testing. The software aims to establish a baseline for DNA synthesis screening, addressing three key challenges:

1. Reducing costs and meeting commercial needs for synthesis providers
   - Cheaper screening processes are necessary as synthesis costs decrease and order volumes increase, creating an economic burden for providers.
   - Providers prefer in-house screening to protect sensitive customer data and maintain trust.
2. Building international trust in screening practices
   - There is limited international collaboration on screening standards, with few shared practices.
   - U.S. resources for screening are tied to national defense and not widely accessible or trusted globally.
3. Integrating screening capabilities into benchtop devices
   - As benchtop synthesis devices become more common, integrating effective and efficient screening into these devices is necessary.
   - Many manufacturers plan to use a "phone home" approach for cloud-based screening, raising similar challenges to traditional providers.

The Common Mechanism screens sequences $\geq$ 50 nucleotides and flags sequences of concern using three analysis modules:

1. **Biorisk Database Comparison (M1)**: Uses hidden Markov Models models to detect variants using an initial database from public sources limited to sequences in regulated pathogens and toxins.
2. **Taxonomic Best Match (M2)**: Compares orders against public DNA/protein sequences and cross-references with international control lists using BLAST and/or DIAMOND.
3. **Benign Gene Identification (M3)**: Evaluates sequences matching regulated, non-viral organisms and identifies known benign functions.

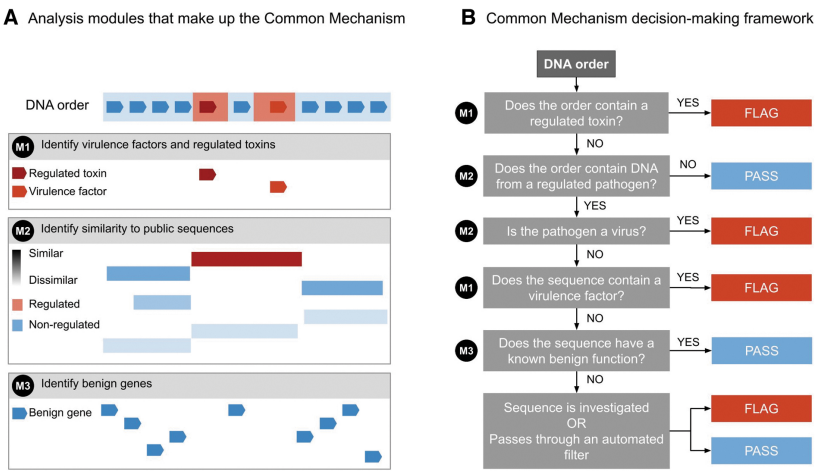


Figure 4: Common Mechanism A) analysis module and B) decision-making framework (from Wheeler, 2024)

Figure 4B shows the decision-making framework for the Common Mechanism, which flags all regulated pathogen and toxins, as well as virulence factors from non-viral regulated pathogens consistent with export controls.

3.3 Testing

To evaluate the Common Mechanism, I installed the software on my laptop and tested it using the SARS-CoV-2 Nucleoprotein sequence, which I had previously ordered from a DNA synthesis company as part of my PhD research. I ran the software in `--fast` mode, which only requires the smaller biorisk and benign databases, rather than the full NCBI databases that demand 275 - 650 GB of disk space.

The screening process successfully identified sequence similarity to the SARS-realted Nucleoprotein:

```
commec screen -d dbs/ -o ./fast-cov example_data/cov.fasta --fast
```

```
## >> STEP 1: Checking for biorisk genes...
##       --> Biorisks: Regulated genes not found, PASS
##       --> Virulence factor found in bases 452 to 870, WARNING
##           Gene: nucleocapsid protein [Severe acute respiratory syndrome-related coronavirus]
##  STEP 1 completed at 2024-07-26 13:45:24
##  >> FAST MODE: Skipping steps 2-3
##  SKIPPING STEP 3: Nucleotide search
## >> STEP 4: Checking any pathogen regions for benign components...
##  ...no regulated regions to clear
## >> COMPLETED AT 2024-07-26 13:45:26
```

This was then flagged as a virulence factor:

```
commec flag .
```

query	biorisk	vf	reg_virus	reg_bacteria	reg_euk	mix_reg_non_reg	benign
SARS-2_N	P	F	-	-	-	-	-

4 Improving the Common Mechanism

4.1 Expanding the Common Mechanism

Wheeler et al. 2024 suggested three areas for improving the Common Mechanism:

Suggestion	Example(s)
1. Expand beyond regulated pathogens and toxins	Expand the list of sequences of concern to include those not derived from regulated pathogens and toxins, as stated in the updated HHS Screening Framework.
2. Update screening as science and policy develop	IBBIS could host and update sequence databases as new sequences of concern are identified or policies change. Respond to advances in deep learning methods that could evade screening or better identify sequences of concern.
3. Limit information hazards from shared sequence databases	Consistent with the HHS Screening Framework, only share databases containing novel sequences of concern with trusted partners.

Table 3: Suggestions for improving the Common Mechanism (Wheeler et al. 2024).

4.2 My Suggestion for Improvement

Building upon the second suggestion, I propose **developing and incorporating an additional analysis module into the Common Mechanism to detect AI-generated sequences.**

4.2.1 Background

AI models, such as those listed in Table 4, are increasingly used to design functional proteins with low sequence similarity to naturally occurring proteins (Wheeler et al. 2024). This presents a challenge for DNA synthesis screening, as AI-generated sequences may evade existing screening methods.

Model	Architecture	Modalities
ESM-3	Transformer	Sequence, structure, function
RF Diffusion	Diffusion	Sequence, structure
EvoDiff	Diffusion	Sequence

Table 4: Examples of widely used AI models for protein design.

The detection of AI-generated protein sequences, remains a nascent field with limited direct research available. Currently, we do not know if AI-generated sequences can be detected reliably. Here, I propose that methods from other domains, such as large language model (LLM) text detection, could be applied to this problem and the resulting method could be integrated into the Common Mechanism if successful.

4.2.2 Methods

Detecting AI-generated sequences is essentially a binary classification task. Perhaps the most reliable current method for detecting AI-generated text is watermarking, where a unique signature is embedded in the generated text during the generation process (Kumarage et al. 2024). However, as this requires cooperation from AI developers, which may not always be feasible, I propose evaluating post-hoc detection methods based on the following two approaches:

1. Feature-based approaches:
    - Use a large dataset of labelled AI-generated and naturally occurring proteins to train a classifier to distinguish between the two.
    - Learn features such as amino acid preferences, sequence motifs, or structural characteristics.
    - Evaluate traditional machine learning algorithms and deep learning models, adapting techniques like TriFuseNet for protein sequences.
2. LLM probability function approaches:
    - Employ techniques like DetectGPT by introducing small changes to protein sequences and compare log probabilities using a pre-trained protein language model.
    - This allows for the detection of AI-generated sequences without extensive training data by directly evaluating probability scores of original and perturbed sequences.

However, many challenges must be addressed:

1. **Feasibility:** Determining the extent to which AI-generated and natural sequences can be distinguished (see Figure 5).
2. **Generalisability:** Ensuring the method works across different AI models
3. **Accuracy:** Minimising false positives and negatives
4. **Adaptability:** Keeping pace with new AI models
5. **Computational demands:** Managing the high computational cost of LLM inference and running this additional module

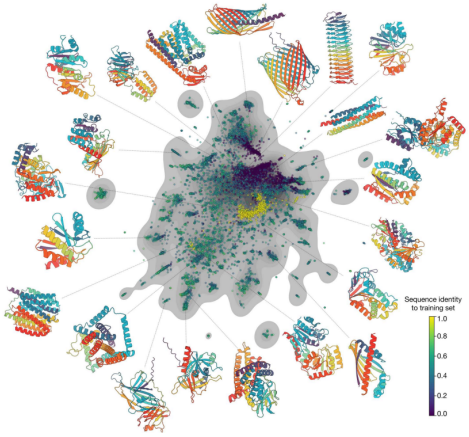


Figure 5: UMAP plot of ESM-3 generated sequences (colored by sequence identity) and natural sequences (gray) (Figure from Hayes et al. 2024).

As shown in Figure 6, the AI detection module could be an additional step after the pathogen detection. This may help prevent AI-dependent evasion of screening, as sequences flagged as AI-generated would be subject to additional scrutiny within the decision-making framework and/or manual investigation.

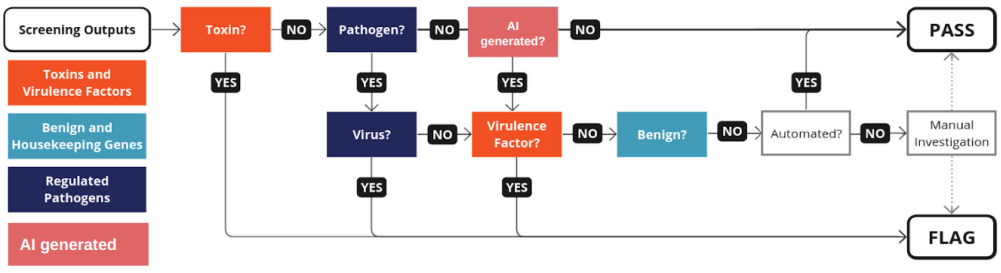


Figure 6: Integration of AI detection module into the Common Mechanism (adapted from IBBIS 2024).

4.2.3 Discussion

Reliable detection of AI-generated protein sequences would be valuable not only for DNA synthesis screening but also for other biosecurity applications, such as early pathogen detection and genetic engineering detection. By addressing these challenges and integrating AI detection into the Common Mechanism, we can enhance our ability to identify potential biosecurity risks and stay ahead of emerging threats in synthetic biology.

References

Carlson, Rob. 2022. “DNA Cost and Productivity Data, Aka "Carlson Curves".” *Synthesis*. http://www.synthesis.cc/synthesis/2022/10/dna-synthesis-cost-data.

Delaney, Oscar, and Hanna Pálya. 2022. “Airtable of DNA Synthesis Companies Globally.” https://airtable.com/appexQt9cKcydq2mm/shrjN5YiVBS464w9s/tblZt70y1is0OP7w1.

Frank, Toby. 2024. “Evaluation of DNA Synthesis Screening Methods.” *Google Docs*. https://docs.google.com/document/d/1afStBicw3XJscBS-nl9q264eqSPS19aoFiAE5a5SBoE/edit.

Gerstein, Daniel M., Bianca Espinosa, and Erin N. Leidy. 2024. “Emerging Technology and Risk Analysis: Synthetic Pandemics.” RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2882-1.html.

Herfst, Sander, Eefje J. A. Schrauwen, Martin Linster, Salin Chutinimitkul, Emmie de Wit, Vincent J. Munster, Erin M. Sorrell, et al. 2012. “Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets.” *Science (New York, N.Y.)* 336 (6088): 1534–41. https://doi.org/10.1126/science.1213362.

Kumarage, Tharindu, Garima Agrawal, Paras Sheth, Raha Moraffah, Aman Chadha, Joshua Garland, and Huan Liu. 2024. “A Survey of AI-Generated Text Forensic Systems: Detection, Attribution, and Characterization.” arXiv. http://arxiv.org/abs/2403.01152.

Noyce, Ryan S., Seth Lederman, and David H. Evans. 2018. “Construction of an Infectious Horsepox Virus Vaccine from Chemically Synthesized DNA Fragments.” *PLoS ONE* 13 (1): e0188453. https://doi.org/10.1371/journal.pone.0188453.

Sugishima, Masaaki. 2003. “Aum Shinrikyo and the Japanese Law on Bioterrorism.” *Prehospital and Disaster Medicine* 18 (3): 179–83. https://doi.org/10.1017/s1049023x00001023.

Thadani, Nicole N., Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Daniel Ritter, Chris Sander, Yarin Gal, and Debora S. Marks. 2023. “Learning from Prepandemic Data to Forecast Viral Escape.” *Nature* 622 (7984): 818–25. https://doi.org/10.1038/s41586-023-06617-0.

Thi Nhu Thao, Tran, Fabien Labroussaa, Nadine Ebert, Philip V’kovski, Hanspeter Stalder, Jasmine Portmann, Jenna Kelly, et al. 2020. “Rapid Reconstruction of SARS-CoV-2 Using a Synthetic Genomics Platform.” *Nature* 582 (7813): 561–65. https://doi.org/10.1038/s41586-020-2294-9.

Wheeler, Nicole E., Sarah R. Carter, Tessa Alexanian, Christopher Isaac, Jaime Yassif, and Piers Millet. 2024. “Developing a Common Global Baseline for Nucleic Acid Synthesis Screening.” *Applied Biosafety* 29 (2): 71–78. https://doi.org/10.1089/apb.2023.0034.

Williams, Bridget, and Rowan Kane. 2023. “Preventing the Misuse of DNA Synthesis.” *Institute for Progress*. https://ifp.org/preventing-the-misuse-of-dna-synthesis/.