U.S. Department of Health and Human Services

ASPR Administration for Strategic Preparedness & Response

Screening Framework Guidance

Matt Sharkey, Ph.D., Biologist

Division of Strategy, Office of Strategy, Policy and Requirements

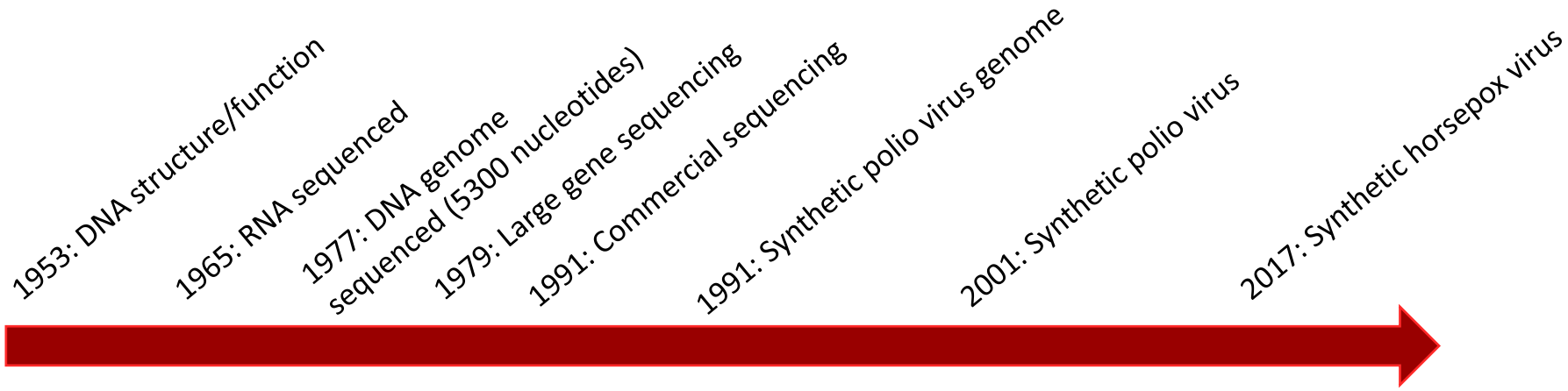
Biosecurity Sequence Screening Training Course for Bioengineers

March 31, 2023

The views, opinions, and/or findings expressed are those of the author and should not be interpreted as representing the official views, policies, or intentions of the U.S. Department of Health and Human Services or the U.S. Government.

Biological Safety Policy Challenge: *De novo* virus synthesis

- Cells can produce viruses if viral genomes are introduced into them
- Prior to the past 20 years, viral genomes were too large to synthesize
  - Polio virus : 7,500 nucleotides of ssRNA – 1991
  - Horsepox virus : 212,000 nucleotides of dsDNA – 2017
- Chemical synthesis challenge is evaporating

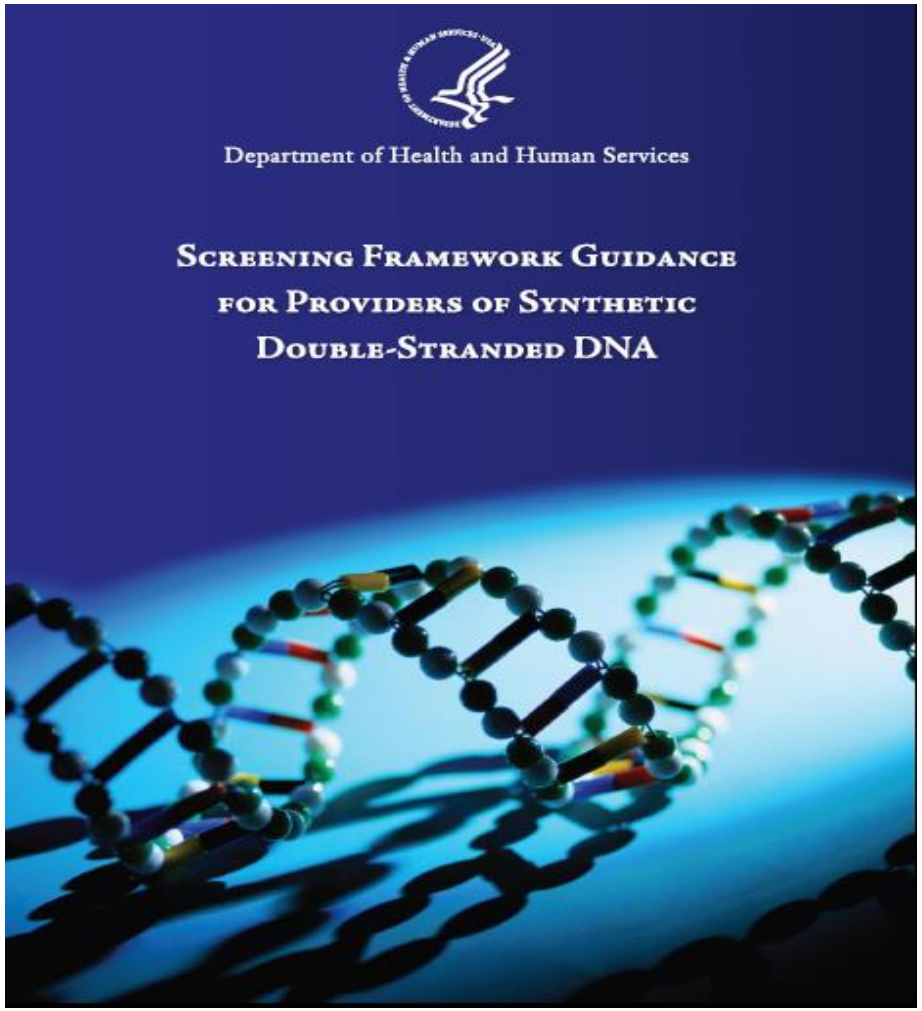


2010: Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA

Primary Goals:

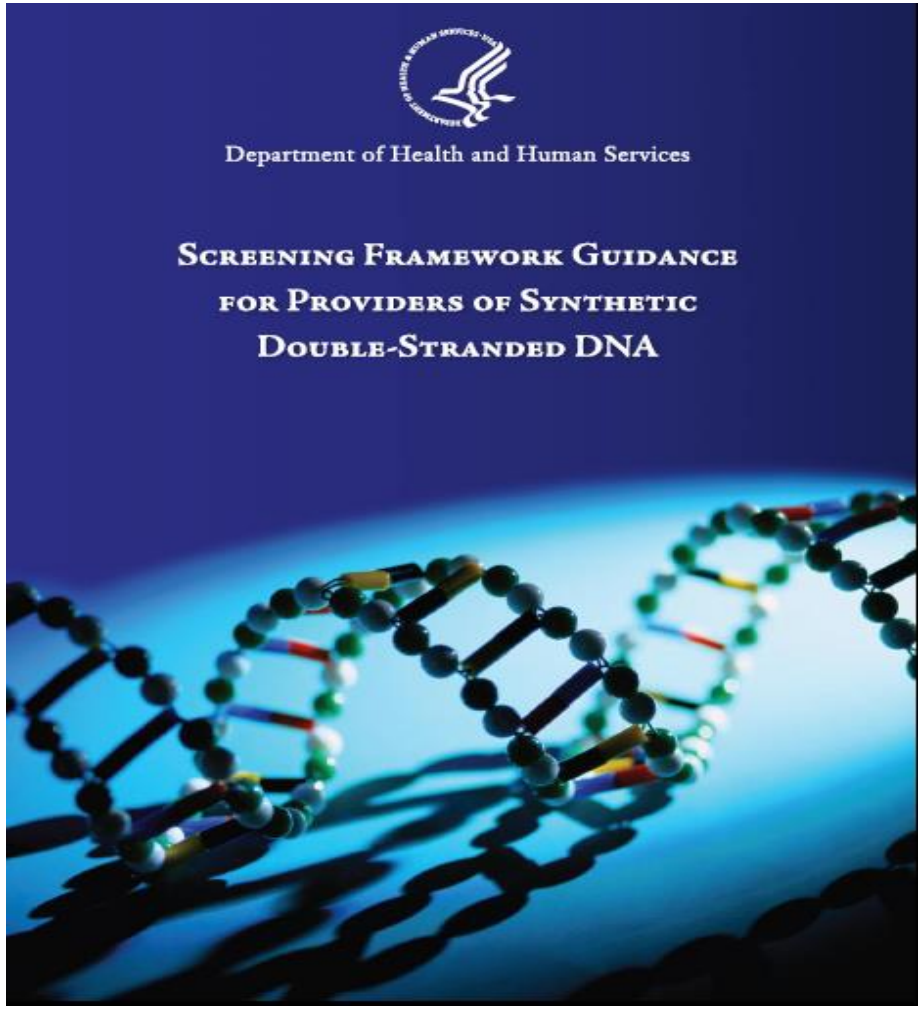
- Reduce the risk that unauthorized individuals or individuals with malicious intent will obtain regulated agents (FSAP or CCL) through the use of commercial dsDNA synthesis
- Minimize any negative impacts on the conduct of research and business operations

- https://aspr.hhs.gov/legal/syndna



Key Elements in 2010 Guidance

- dsDNA synthesis providers are the target of the Guidance
- Encourages screening 200 bp windows for BSAT & CCL sequences
- Voluntary Guidance, resulted in wide uptake (e.g., IGSC adoption)
- Encourages providers to know customers and whether products they are selling pose a hazard
- Including guidance for contacting USG if malintent is suspected



The International Gene Synthesis Consortium (IGSC)

- The International Gene Synthesis Consortium (IGSC) is an industry-led group of gene synthesis companies and organizations formed to design and apply a common protocol to screen both the sequences of synthetic gene orders and the customers who place them.

- Currently, the IGSC members together represent approximately 80% of commercial gene synthesis capacity world-wide.

  - **ATUM (formerly DNA2.0)**
  - **Battelle**
  - **BGI**
  - **Bioneer Corp**
  - **Codex DNA**
  - **Blue Heron Biotech**
  - **The DAMP lab**
  - **Codex DNA**
  - **Edinburgh Genome Foundry**
  - **Evonetix**
  - **GenScript USA**
  - **Ginkgo Bioworks**
  - **IDT**
  - **Nuclera**
  - **Raytheon BBN Technologies**
  - **Thermo Fisher Scientific**
  - **Twist Bioscience**

California and Maryland state legislatures consider bills to require compliance with IGSC or HHS Screening Frameworks

- CA Assembly Bill 70 passes in September 2021
  - Gov Gavin Newsome vetoed
- MD House Bill 1256 Introduced February 8, 2021, withdrawn April 2021

Bills each require entities in receipt of state funds to only purchase synthetic genes from companies who are either IGSC verified or compliant with HHS SynDNA Guidance

Bills each require companies synthesizing genes to be compliant with either IGSC or HHS SynDNA Guidance

See: West R, Gronvall GK. California shows the way for biosecurity in commercial gene synthesis. Nat Biotechnol. 2020 Sep;38(9):1021. doi: 10.1038/s41587-020-0667-0. PMID: 32811990.

AB 70, Salas. Gene synthesis providers.

Existing law requires the State Department of Public Health to establish an advisory committee to advise the Legislature and the Governor on human cloning and other issues relating to human biotechnology.

This bill would require the department to develop a process, with input from the International Gene Synthesis Consortium (IGSC) and industry stakeholders, to verify that gene synthesis providers and manufacturers of gene synthesis equipment adhere to customer and sequence screening protocols that are equivalent to, or stronger than, the IGSC's Harmonized Screening Protocol. Beginning January 1, 2025, the bill would require a gene synthesis provider and manufacturer of gene synthesis equipment operating in California to be a current member of the IGSC or verified by the department as adhering to the prescribed proper screening protocols. The bill would, beginning January 1, 2022, authorize the department to charge a fee in an amount not to exceed the department's reasonable costs to establish and administer the verification process, as specified. The bill would also require, beginning January 1, 2025, any entity that is the recipient of state resources to purchase gene synthesis products from a gene synthesis provider, and gene synthesis equipment from a manufacturer of gene synthesis equipment, only if they are a current member of the IGSC or verified by the department, as specified. The bill would impose specified penalties on gene synthesis

Biological Safety Policy Challenge: *De novo* virus synthesis

How Canadian researchers reconstituted an extinct poxvirus for $100,000 using mail-order DNA

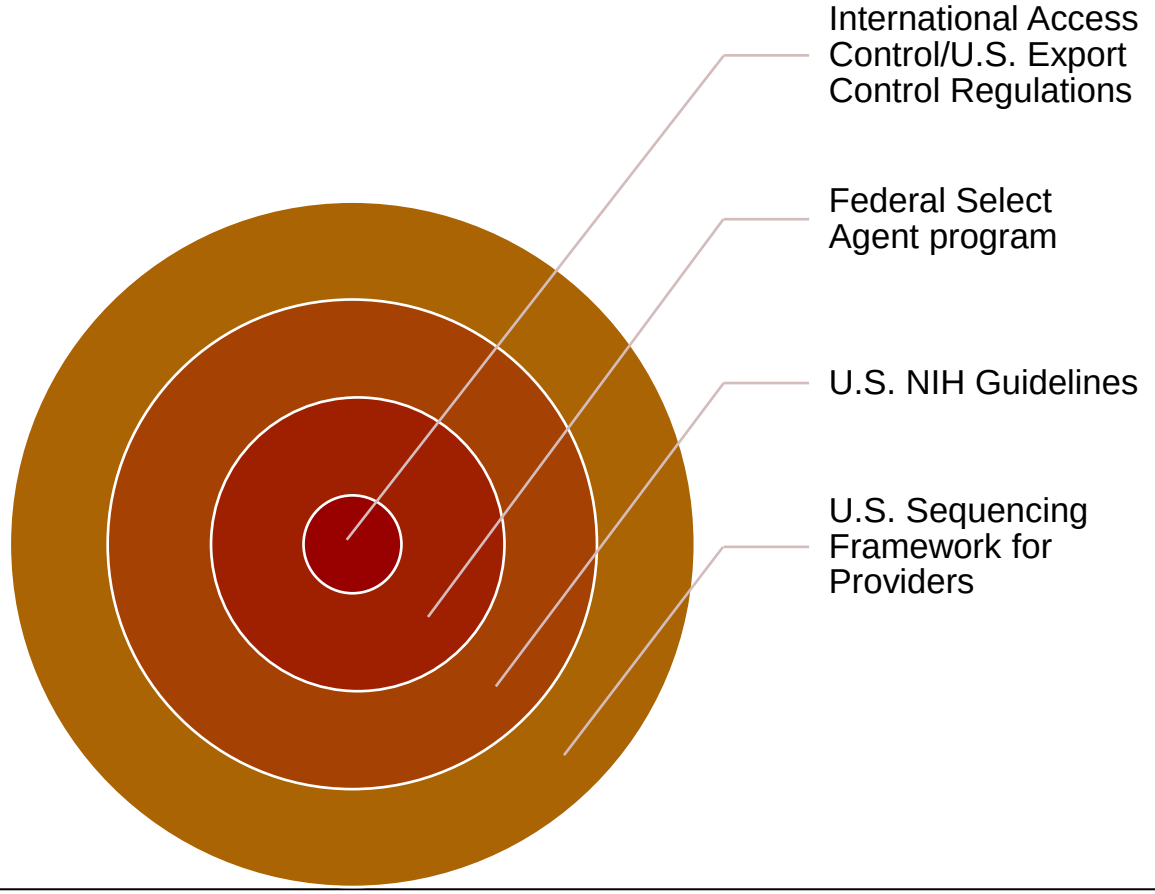
By Kai Kupferschmidt | Jul. 6, 2017 , 5:00 PM

http://www.sciencemag.org/news/2017/07/how-canadian-researchers-reconstituted-extinct-poxvirus-100000-using-mail-order-dna

Horsepox virus DNA sequence is more than 95% similar to smallpox (variola)

- Many 200 bp windows are shared between horsepox and smallpox viruses

Existing in only two laboratories, in Novosibirsk and Atlanta, smallpox virus is the most access-restricted material that exists. The technology for producing it *de novo* has been published and the act raised no red flags. Biological safety policies need to be updated.



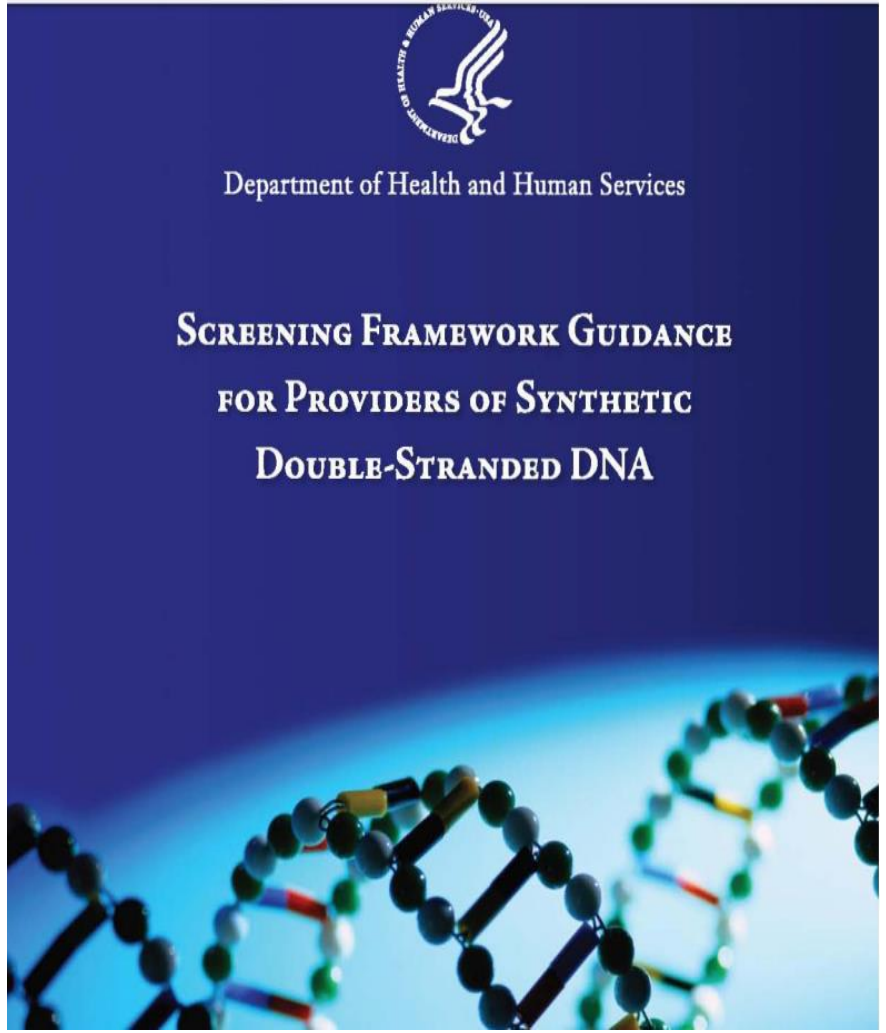
Noyce, R.S., Lederman, S., Evans, D.H. (2018). Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments. *PLOS ONE, 13*(1): e0188453

Biological Safety Policy Challenge – Evolving Biotech Landscape

- Advances in technology
  - Increasing ease of large-scale assembly.
  - Synthesis of horsepox virus from synthetic oligonucleotides in 2017
  - Ease of conversion between types of oligonucleotides
- Increasingly available and affordable
  - Cost per base pair: $0.05 - $0.17, with new methods promising lower costs
- Privatized
  - 45 companies in 2010, more than 320 companies in 2022
- Globalized
  - Global value of the DNA synthesis market grew from $46.3 million in 2010 to $203 million in 2017, projected growth to 2023 up to $825 million.
- Democratized
  - Readily purchase reagents and automated equipment on-line (e.g., benchtop synthesizers)
  - Demographics of practitioners includes scientists and others (e.g., DIY community)

Biological Safety Policy Challenges: Reestablishing Access Control

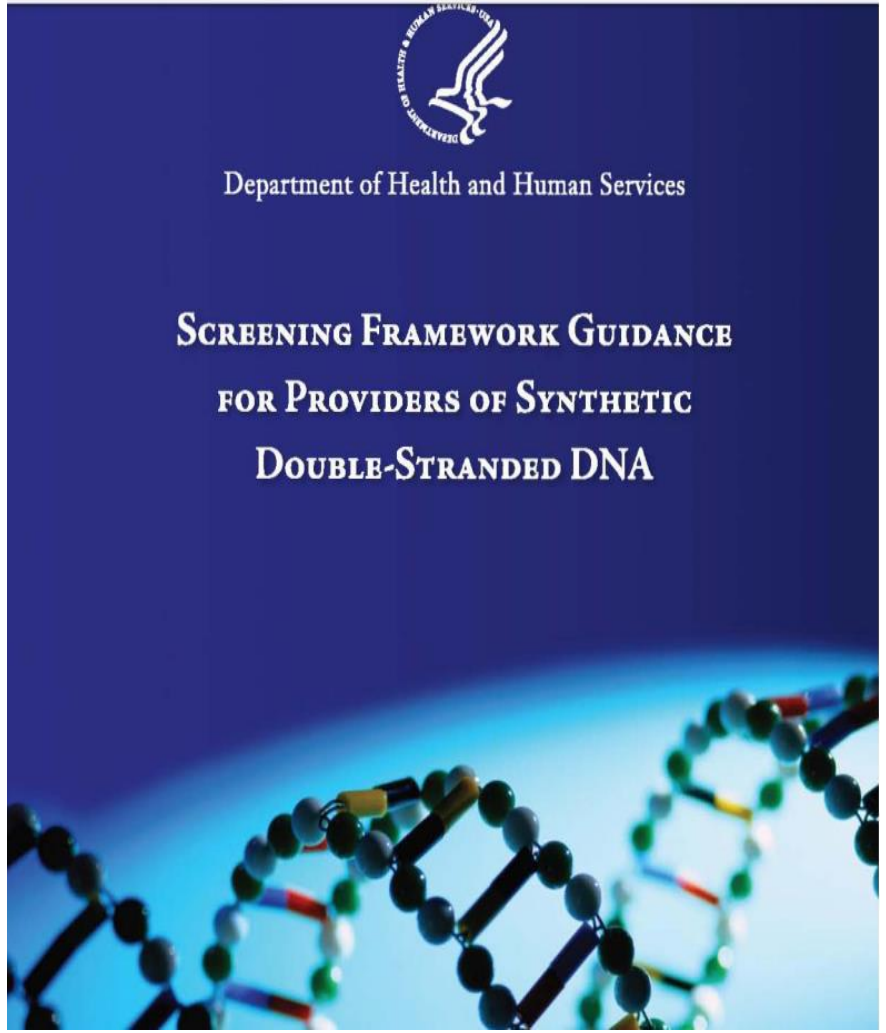
- Screening Framework Guidance for Providers of Synthetic Double-stranded DNA (2010) has been reviewed for potential updates since 2017
- Two Federal Register Notices soliciting stakeholder comments, published in 2020 and 2022
  - 2020: Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA: https://shorturl.at/jozJS
  - 2022: Screening Framework Guidance for Providers and Users of Synthetic Nucleic Acids: https://shorturl.at/cnzE6



Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA
FRN 2020-18444 August 2020 – January 4, 2021

Stakeholder (industry, academia, government) input was requested on:

- Scope of the Guidance
- Sequence Screening
- Biosecurity Measures
- Customer Screening
- Minimizing Burden of the Guidance
- Technologies Subject to the Guidance
- Other considerations



Screening Framework Guidance for Providers and Users of Synthetic Oligonucleotides – FRN 2022-09210, April 29, 2022 – June 28, 2022

2010 Guidance	2023 Guidance
Scope	
Providers	Providers, Users , Manufacturers,* Third-party Vendors, Institutions, Principal Users, End Users *of benchtop nucleic acid synthesizers
Sequences of Concern (SOCs)	
Limited to those from agents on Biological Select Agents and Toxins list & Commerce Control List	Sequences of Concern definition, to include all sequences that may confer pathogenicity or toxicity
Oligonucleotide Type	
Double-stranded DNA	Single- and double-stranded DNA and RNA
Sequence Screening Window	
200 bp	50 bp; <50 bp for batch orders with the potential for assembly into SOCs
Records Management & Cybersecurity	
Records Management	Records management (e.g., Material Transfer Agreements), Cybersecurity

Sequence of Concern Definition – Expect Difficulties Implementing

Sequence of Concern (SOC): A nucleotide sequence that is a Best Match (see the SEQUENCE SCREENING METHODOLOGY section of this Guidance) to a sequence of federally regulated agents (i.e., the Biological Select Agents and Toxins List, or the CCL), except when the sequence is also found in an unregulated organism or toxin. As soon as it is practical to do so, it is also recommended that sequences known to contribute to pathogenicity or toxicity, even when not derived from or encoding regulated biological agents, be treated as SOCs. Follow-up customer screening to verify legitimacy should take place when a SOC is identified (see Verifying Legitimacy in this section of the Guidance). [8, 9]

[8] Organizations should define and document their criteria for determining whether a sequence is of concern. In order to ensure compliance with the FSAP and CCL regulations, sequences of concern should include sequences derived from their listed agents – except when they are also found in unregulated agents.

[9] Pathogenicity or toxicity that threatens public health, agriculture, plants, animals, animal or plant products, or the environment. SOCs include sequences for which a direct and harmful impact on a host has been verified based on published experimental data; and, where experimental data do not exist, based on homology to a sequence encoding a verified function.

Sequence of Concern Definition – Expect Difficulties Implementing

Sequence of Concern (SOC): A nucleotide sequence that is a Best Match (see the SEQUENCE SCREENING METHODOLOGY section of this Guidance) to a sequence of federally regulated agents (i.e., the Biological Select Agents and Toxins List, or the CCL), except when the sequence is also found in an unregulated organism or toxin. As soon as it is practical to do so, it is also recommended that sequences known to contribute to pathogenicity or toxicity, even when not derived from or encoding regulated biological agents, be treated as SOCs. Follow-up customer screening to verify legitimacy should take place when a SOC is identified (see Verifying Legitimacy in this section of the Guidance). [8, 9]

[8] Organizations should define and document their criteria for determining whether a sequence is of concern. In order to ensure compliance with the FSAP and CCL regulations, sequences of concern should include sequences derived from their listed agents – except when they are also found in unregulated agents.

[9] Pathogenicity or toxicity that threatens public health, agriculture, plants, animals, animal or plant products, or the environment. SOCs include sequences for which a direct and harmful impact on a host has been verified based on published experimental data; and, where experimental data do not exist, based on homology to a sequence encoding a verified function.

Addressing Risks Associated with Benchtop Synthesizers

- Manufacturers should screen all Customers purchasing benchtop nucleic acid synthesizers to validate customer legitimacy and that the equipment is appropriate for their needs. Manufacturers should only provide nucleic acid synthesizers to Customers that have mechanisms in place that aim to ensure that the devices are only operated by legitimate users.
- **What about equipment sold prior to the Guidance?**
- **Is there an Institutional responsibility to ensure only legitimate users access benchtop synthesizers?**
- **Should benchtop synthesizers screen against a hardwired SOC database or use cloud screening?**

Addressing Risks Associated with Unknown Sequences

Although there are likely legitimate explanations for orders of sequences with no matches in existing databases (e.g. nucleic acids ordered to populate microarrays or to store digital information), in such cases, **it may be possible to use predictive bioinformatic algorithms to screen sequences that are not a match to any known sequences to determine if they could produce proteins that are structurally and functionally identical to SOCs**. This Guidance encourages Providers to continue to develop these methods to best ensure the safety and security of the synthetic nucleic acid research enterprise.

Addressing Risks Associated with Unknown Sequences

Urbina, Fabio, Filippa Lentzos, Cédric Invernizzi and Sean Ekins. 2022. Dual use of artificial-intelligence-powered drug discovery. Nature Machine Intelligence 4, Nr. 3: 189–191. doi:10.1038/s42256-022-00465-9

- Elkins, S., Lentzos F, Brackmann M, Invernizzi C. 2023. There's a 'ChatGPT' for Biology. What Could Go Wrong? Bulletin of the Atomic Scientists. March 24. https://thebulletin.org/2023/03/chat-gpt-for-biology/
- Ferruz, N., Schmidt, S. & Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. Nat Commun 13, 4348 (2022). https://doi.org/10.1038/s41467-022-32007-7
  - *These results have impactful consequences for protein engineering because ProtGPT2 appears to preserve binding positions in the generated sequences, despite the low identities (31.1 and 29.2% for 357 and 45, respectively)…*
- Madani, A., Krause, B., Greene, E.R. *et al.* Large language models generate functional protein sequences across diverse families. *Nat Biotechnol* (2023). https://doi.org/10.1038/s41587-022-01618-2
  - *Artificial proteins fine-tuned to five distinct lysozyme families showed similar catalytic efficiencies as natural lysozymes, with sequence identity to natural proteins as low as 31.4%...*

Matt Sharkey, Ph.D.
Email: matthew.sharkey@hhs.gov

On the Web:
aspr.hhs.gov

Facebook:
facebook.com/ASPRgov

Twitter:
twitter.com/ASPRgov

Twitter: Dawn O'Connell
twitter.com/HHS_ASPR

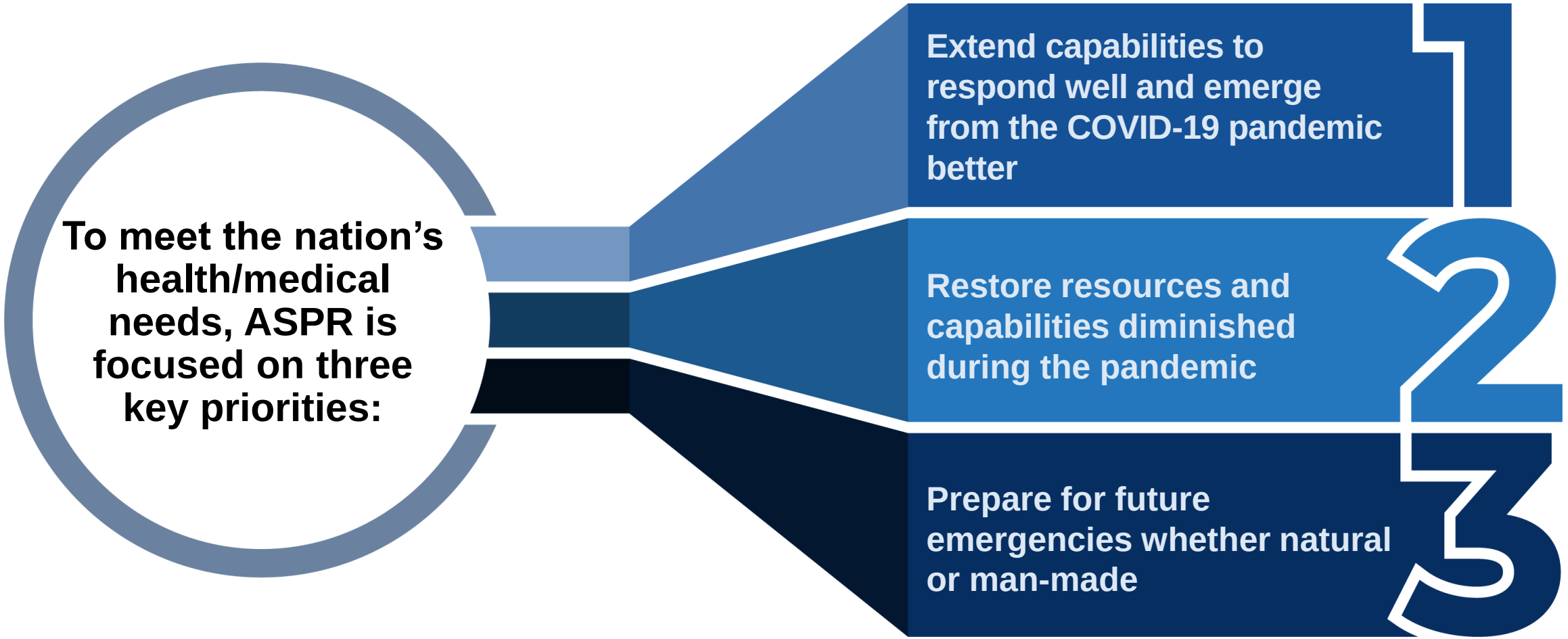
Instagram:
instagram.com/ASPRgov/

YouTube::
youtube.com/c/ASPRgov

Flickr:
flickr.com/ASPRgov

LinkedIn:
linkedin.com/showcase/hhs-aspr/

ASPR Key Priorities



Disclaimer

The views, opinions, and/or findings expressed are those of the author and should not be interpreted as representing the official views, policies, or intentions of the U.S. Department of Health and Human Services or the U.S. Government.