

## Comments Received in Response to Federal Register Notice 2020-18444, Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA

Rapid and continued advances in nucleic acid synthesis technologies and synthetic biology applications necessitate periodic reevaluation of associated risks and appropriate mitigation measures. To determine how risk mitigation measures should be balanced with the need to support both scientific progress and the success of the U.S. biotechnology enterprise, the Department of Health and Human Services' Office of the Assistant Secretary for Preparedness and Response issued the Request for Information for the Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA in the U.S. Federal Register, on August 26, 2020. The public was invited to submit comments on whether and how the Guidance could be updated to mitigate the emerging risks associated with nucleic acid synthesis technologies.

More specifically, input on potential changes that would either expand or limit the following topical areas was sought:

- Scope of the Guidance;
- Sequence Screening;
- Biosecurity Measures;
- Customer Screening;
- Minimizing Burden of the Guidance; and
- Technologies Subject to the Guidance.

Comments received in response to this Federal Register publication were considered by the U.S. government in reviewing the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA and composing the Request for Information on the Screening Framework Guidance for Providers and Users of Synthetic Oligonucleotides. Those comments are included in this document, with personally identifiable information redacted. They are arranged by questions associated with each of the topics.

1-1

**Scope of the Guidance - Question 1: Should the focus of the Guidance Extend beyond the Select Agents and Toxins list and CCL?**

---

The select agent and toxins list and the CCL are not comprehensive or agile enough to address the threat.

Submitted on: 9/15/2020 9:16:26 AM

Agency Type: Federal Government / Agency Other:

---

## 1-2

Yes. There are two fundamental issues with the current Guidance: (1) there is no homogenization on the definition for the fundamental unit of control and (2) there is no rule-based criteria for database selection/curation and flagging for review. The latter issue will be addressed in our response to the best match criteria question, and the former issue will be addressed here. The former issue has resulted in a discrepancy between the products of synthetic dsDNA providers (genes/gene fragments) and what is controlled (typically organisms). Even in the case of toxins the current Guidance reference several lists that control sequences at different biological levels. For example, the Department of Commerce describes a gene as the fundamental unit of control, whereas the Commerce Control List describes a protein chain as the fundamental unit of control. This lack of homogenization in the definition of the fundamental unit of control results in confusion among DNA screening practitioners (e.g., the International Gene Synthesis Consortium (IGSC) recently approved an advisory opinion request to the Department of Commerce seeking clarification on the department's definition of a gene for purposes of control) and creates unnecessary technical challenges and fallacies for the detection of potentially controlled sequences in dsDNA orders. Thus, a clear definition of the fundamental unit of control is the first major step to improving the current Guidance and will ultimately determine the scope of the Guidance.

Given our decades of experience studying and mitigating biothreats, as well as our more recent work not only in providing screening services to dsDNA providers, but also in building next generation bioinformatic tools, we suggest that the fundamental unit of control should be at the level of biothreat function. Homogenizing the fundamental unit of control to this definition would result in both a contraction in the current scope of the Guidance, as the majority of non-threatening genes in controlled organisms will no longer be controlled, and an expansion of the current scope of the Guidance, as sequences that result in biothreat function can be found in organisms not just those found on select lists. The major benefit of such a definition is that it is specific enough to clearly define a biothreat, while staying broad enough to cover emerging biothreats, so that the definition does not suffer from the same issues plaguing the Chemical Weapons Convention (CWC) (e.g., the development of Novichok agents, which eluded the CWC until November of 2019). Further, this fundamental unit of control is in sync with the products of synthetic dsDNA providers. It is important to note that this unit of control does not necessarily need to be a full gene. For example, the enzymatic toxic activity associated with of Ricin A chain (Uniprot accession P02879) is encoded in amino acids 36-302 (of a 576 amino acid protein). Similarly, the NS1 chain of Yellow Fever Virus, which is responsible for immune evasion of the virus (Uniprot accession Q6DV88) is encoded in amino acids 779-1130 (of a 3,411 amino acid protein).

Biothreat functions are encoded by sequences of concern (SoCs) and are associated with pathogenicity, toxicity, drug production, and other functions or end products that can be detrimental to high priority hosts. While humans are an obvious high priority host, we suggest that non-human hosts of high economic value should also be considered as high priority hosts. More specifically, an analysis performed by the United States Department of Agriculture Economic Research demonstrated that cattle, poultry, and swine comprised 96% of U.S. livestock farm receipts (of \$176 billion) and corn,

soybeans, and wheat comprised 48% of U.S. crop farm receipts (of 195.4 billion) in 2017. Together, these six commodities comprised 71% of all U.S. farm receipts in 2017 and thus these hosts should be considered high priority.

Once high priority hosts are defined, SoCs that impact these hosts must be curated in a standardized fashion. We propose the following standardized definitions for biothreat functions that can be used to curate SoCs and a Tiered criterion for linking function to DNA screening review:

Tier 1: SoCs that directly trigger review. SoCs that are known to encode functions that directly cause adverse effects to the host that they impact. These SoCs encode the following biothreat functions:

Damage: Directly damages host cells, cellular processes, or cellular barriers (e.g., Ricin toxin A chain; conotoxins; Andes virus N protein, which leads to capillary leakage).

Active host subversion: Actively aggravates host immune detectors or effectors (e.g., *Staphylococcus aureus* chemotaxis inhibitory protein, which blocks neutrophils and complement-mediated killing; *Yersinia* Yop proteins, which downregulates pro-inflammatory response).

Tier 2: SoCs that trigger review in a specific context. SoCs that encode functions that are known to enhance the pathogenicity of known regulated organisms (these SoCs must be identified within a single sequence). We suggest defining the regulated list of controlled organisms as the combined list of organisms from the following sources, as these organisms were selected based on their unique characteristics that enable the potential for weaponization: Federal Select Agent Program, Commerce Control List, and Australia Group List. These SoCs encode for the following functions:

Antimicrobial resistance: Confers antibiotic / antiviral resistance to a regulated pathogen (e.g., Ampicillin resistant *Yersinia pestis*) [Note: review only triggered when both an antibiotic resistance function and a regulated pathogen are uniquely identified within the same sequence.]

Bioregulator: Enhances regulated pathogens through addition to host cell regulators of high impact human systems such as the cardiovascular, nervous, and immune systems (e.g., Interleukin-4 in combination with a poxvirus)

Tier 3: SoCs that trigger review in aggregate (a quantitative metric for triggering is defined in our response to the best match criteria question). SoCs that trigger review only when multiple SoCs from the

same pathogen are identified within or across orders from a single source. These SoCs encode for the following functions:

Passive host subversion: Avoids immune surveillance by altering recognizable elements of the pathogen, repairing damages caused by the host immune system, or other indirect means (e.g., Ebola virus glycoprotein, which avoids immune surveillance via epitope masking, steric shielding, and decoying)

Inhibits host cell death: Suppresses host cell to allow replication or avoid pathogen defeat (e.g., EspZ protein from *Escherichia coli*, which stalls premature host cell death).

Promotes host cell apoptosis: Activates host cell death to allow replication or void pathogen defeat (e.g., HIV envelop proteins, which induce apoptotic signals).

Adherence/Invasion: Enables a pathogen or toxin to bind to and/or actively enter or maintain protected spaces within the host (e.g., Ricin toxin B chain; invasion plasmid antigen A from *Shigella* sp., which enables invasion through actin dysregulation; PilC and PilE adherence proteins from *Neisseria meningitidis*)

Motility: Enables a pathogen to move within or between host cells (e.g., ActA from *Listeria monocytogenes*, which activates host actin polymerization machinery to propel the pathogen)

Drug, toxin, and explosive pathway enzymes: Enzymes that uniquely lead to the formation of toxins, drugs, and explosives or precursors thereof of interest (e.g., enzymes that produce aflatoxins, trichothecene mycotoxins, microcystins, tetrodotoxins, saxitoxins, opiates, and cannabinoids)

Further elaboration of these functional categories to provide clear, granular definitions and examples for each function is available upon request.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 1-3

Adding sequences of interest beyond select agents/toxins/CCL will increase the cost of the screening process itself, not only in the up-front computational screening, but also importantly in downstream domain expert follow up screening, and follow up conversations with the user/customer as well as the FBI/counterintelligence personnel. To justify these additional costs, there must be corresponding/proportional benefits (e.g. to reduce the likelihood/severity of a biosecurity risk). While there are clearly sequences that are of concern/pose a threat to biosecurity/biosafety beyond the select agents/toxins/CCL, the challenge will be to do the cost/benefit analysis; especially since the costs and benefits will not be consistent across sequences, forms of nucleic acids, vendors/providers, or jurisdictions/operational contexts. A related cost comes to the adverse impact on downstream research/development/commercialization, which may be further delayed or disrupted. Thus, scientific knowledge generation could be slowed, and there may be negative consequences to the (bio)economy. Short version: yes, but only following a diligent cost/benefit analysis for each prospective additional sequence of interest, not only on the screening process itself, but also on downstream R&D and commercialization efforts. As further discoveries are made on gene function and roles in pathogenicity, it is important that the Select Agents and Toxins and CCL lists are regularly updated to include or omit sequences involved in pathogenicity or those demonstrated not to play roles, respectively.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other

---

## 1-4

Lists of controlled genes, products, and taxonomic units are an effective and low-ambiguity means of bounding the scope of concern about dangerous materials. While list-based security has been criticized as unable to capture the full scope of possible threats, it is nevertheless the best method available at the current state of knowledge. Alternative approaches based on functionality are still too immature to be effectively applied.

The current lists, however, need to be expanded to include additional pathogens, including emerging and extinct viruses. Moreover, such updates should occur on at least an annual basis, with input from the scientific community, in order to be able to capture emergent or newly discovered agents of concern.

To better support ongoing maintenance of screening systems, whenever possible list entries should be linked with unambiguous definitions, such as NCBI taxonomy IDs. Likewise, the version history of the list should be maintained in a publicly accessible location, in order to better support record-keeping of decisions against its changing content over time.

Finally, in addition to select agents and Tier 1 select agents, it would also be useful to include a third (lower) level of threat for "borderline" threat agents that are not controlled but should nevertheless be subject to a higher level of caution in fulfilling orders. This would allow better management of agents that are not controlled but that are potentially of concern due to being adjacent to controlled agents. An example of such an agent would be *Yersinia pseudotuberculosis*, which is a dangerous pathogen closely related to *Yersinia pestis*, but likely not rising to the level of concern of a select agent.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 1-5

On behalf of NAME, I would like to commend the Department of Health and Human Services (HHS) for publishing this request for information (RFI), Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA, as a first step in the process to update a guidance that is critical to advancing innovation within this field while protecting the security of the United States. In recent years, NAME has been advocating for such an effort and we are very pleased to submit these comments for your consideration and offer our assistance as you work to expeditiously finalize the new Guidance.

NAME is a leading and rapidly growing U.S.-based synthetic biology and genomics company that has developed a disruptive DNA synthesis platform to industrialize the engineering of biology. The core of the platform is a proprietary technology that pioneers a new method of manufacturing synthetic DNA by "writing" DNA on a silicon chip. As you are aware, synthetic biology holds much promise in healthcare, environmental science, manufacturing, and more. Earlier this year, during the early weeks of the COVID-19 pandemic, many clinical laboratories experienced challenges in accessing positive control samples from affected patients to validate their tests. As a real-world example of the positive impact of synthetic biological tools, NAME was able to mobilize quickly to develop and manufacture an inert synthetic RNA viral positive control for COVID-19 testing and work with the U.S. Food and Drug Administration (FDA) to enable its use in test protocol validation procedures. Working in partnership with the Agency, we helped to address one of the major barriers to help rapidly expand testing capacity in the U.S. The ability to contribute to the country's response to the pandemic is just one of the countless benefits of this technology. While we recognize its potential for good, we also are aware of the importance of actively working to ensure this technology is not used for inappropriate purposes.

As an industry leader, NAME promotes the appropriate use of dual use technologies by adhering to standards higher than those in the 2010 HHS guidance, applying more robust protocols when screening both the DNA sequence itself and the customer. NAME believes that updating the HHS guidance to reflect current industry best practices will raise the industry standard among all stakeholders to a level that offers greater protections against the exploitation of synthetic DNA providers and products for misuse. NAME has demonstrated that these high screening standards are realistic and align with the financial interests of a business. More importantly, these rigorous screening practices are essential to allowing this industry to flourish, grow the bioeconomy and contribute to the public good. We offer the following recommendations for the Guidance in support of that goal.

Based on our experience with screening practices, we believe the scope of a future Guidance is the most important aspect that should be under consideration, and as such, we make several high priority recommendations: To better reflect current sequence screening practices, move beyond solely relying on lists of organisms by defining pathogenic sequences of concern



Optimal sequence screening from a provider perspective has three unambiguous outcomes: 1) a clear, repeatable answer as to whether a given ordered sequence is 'of concern' and requires detailed follow-up, 2) clarity around the type of concern posed by a sequence and applicable regulatory framework (if any) and 3) a set of next steps to be carried out by the provider.

Regarding the screening of sequences, currently, the Guidance focuses on identifying sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). Screening against FSAP- or CCL-listed organisms generally can be designed to meet all three of these goals. However, we find the current scope of the Guidance to be non-specific and not complete, and we highly recommend shifting the guidance to also focus on individual DNA sequences of concern.

The creation of both of the FSAP and CCL programs predates cost-effective synthesis of gene-length DNA sequence as well as many of the technologies we now refer to as synthetic biology, resulting in gaps in biosecurity. When FSAP was created, 'reducing the risk of misuse of biology' and 'controlling access to specific weaponizable pathogens' were one and the same. This equivalence is no longer true. The lists of organisms outlined in the FSAP and CCL may be thought of as motivating examples but are both non-exhaustive and non-specific. They are non-exhaustive in the sense that there are sequence components outside of FSAP-listed organisms that, alone or in combination, can be misused to cause harm. And non-specific because, for listed bacteria, the vast majority of sequences in an organism's genome do not 'endow or enhance' pathogenicity and so are not subject to export license requirements for most destinations.

NAME recommends that the Guidance be modified to prioritize a focus on the broader goal of reducing the risk of misuse of biology, ideally by focusing on individual sequences (or even discrete biological functions) as units of control rather than species. This, however, is made extremely difficult by the lack of a U.S. government resource defining pathogenicity-linked sequences and providing contextual information describing the role of each sequence in a well-characterized pathogenic process. Absent such a resource, future iterations of the Guidance may have little choice but to continue to rely on the non-exhaustive and non-specific FSAP and CCL lists.

To fill this gap, many providers build and maintain lists of such sequences and other metadata (e.g. NCBI taxon IDs mapping to controlled organisms) for their own proprietary use. While this effort is necessary given that the field has evolved significantly since HHS's Guidance was released, the lack of a standard list of individual sequences has resulted in non-uniform screening across the industry. Lists vary from provider to provider, in part, because of the limited availability of expertise required to identify such sequences, the expense of acquiring and maintaining that expertise, and the ongoing costs associated with curation of these sequences.

Some providers supplement these lists of regulated sequences and taxon IDs with unregulated sequences that may present biosafety risks to staff under BSL1 conditions. It would be valuable for the Guidance to explain the value of expanding screening to include biosafety concerns including the ability to protect the health and safety of synthesis provider employees and to allow the provider to warn customers of specific biosafety concerns associated with ordered sequences. Importantly, the U.S. government should create and maintain a new list of such sequences for use by domestic providers.

A significant first step to aid in defining when a sequence's pathogenicity is of concern is for the Guidance to describe or define when a sequence is of sufficient length or of sufficient homology (to a known, publicly available gene sequence annotated as coming from a listed organism or toxin) to be considered a 'gene'. Not only will this help inform the list to standardize sequence screening practices among providers, but it will ensure that screening practices continue to be robust in the event the list becomes out of date or the provider is presented with a new order not previously considered.

Further, this definition will also be helpful in informing domestic manufacturers when they are subject to an export license requirement. The language of Export Control Classification Number (ECCN) 1C353 declares the unit of control as 'gene or genes specific to' viruses or bacteria listed in ECCNs 1C351 and 1C354. ECCN 1C353 does not, however, define what constitutes a 'gene' for the purposes of control under ECCN 1C353. Hence, the burden falls on each individual DNA synthesis company to identify sequences of concerns and, when unclear if a license is required, to submit formal requests for item classification to the Department of Commerce.

This current practice results in: (1) increased variability in export compliance by U.S. based DNA synthesis companies, (2) extended turn-around time for sequence delivery, and (3) increased internal labor costs per base pair for U.S. based DNA synthesis companies. These in turn reduce the competitiveness of American synthesis providers globally. Further, the lack of consistent application of such classification across providers leads to risk in misuse. We believe the Department of Health and Human Services (HHS) is the most appropriate federal department to establish what constitutes a gene for the purposes of biosecurity, and we strongly urge HHS to work with its counterparts at the Department of Commerce to streamline and utilize the same definition of 'gene' and sequence that can 'endow or enhance' pathogenicity.

As you consider updating the Guidance to include this definition, specifically, the definition of a 'gene' of concern for this purpose should include all of the following:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

However, it should not include the following:

1. A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.
2. A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Furthermore, the Department of Commerce should align with HHS's definitions for the purpose of providing guidance on whether an export license is required or not. Additionally, if HHS is unable to shift to this approach (i.e. sequence screening) within the Guidance, then it should make clear that novel risks may emerge from collections of sequences outside of traditional weaponizable agents and recommend that DNA synthesis providers should include consideration of these risks in any comprehensive biosecurity program.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-6

Yes, several additional agents could be added, in particular, a short list of carefully selected viruses and newly discovered toxins. For example, poliovirus and MERS are not select agents and can be readily synthesized. A NASEM committee or an organization such as the Global Virus Network would be ideal to help governing bodies choose what viruses and viral strains should be added to the list. Additionally, there are and/or will likely be newly discovered toxins that should be added to the list as they are characterized.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

**1-7**

Several members of our company have discussed these topics and the input represents the entire group.

Yes. We feel that is imperative that we move beyond the limitations of an organism-based list, redefining our focus to that of known gene-based mechanisms (of virulence, toxicity, host immune evasion antibiotic resistance, etc.)

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 1-8

Yes, the Guidance should extend beyond the Select Agents and Toxins list and CCL. Flagging all genes within microbial genomes on the Select Agents and Toxins list and CCL would include tens of thousands of innocuous sequences shared with non-pathogenic, and even nonsymbiotic, phylogenetically close relatives (false positives). The existing guidance also neglects many sequences from disease-causing microbes or toxins that can cause damage and enable infection in humans and economically important species, even though these microbes or toxins may not be weaponizable enough to merit inclusion on select agent lists (false negatives).

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 1-9

Yes, we understand that there are sequences of concern that are not on the lists. The lists could be augmented with sequences from e.g. the (DHS-supported) SOI database at LLNL and Batelle's ThreatSEQ database. However, growing the lists will increase the screening burden on DNA synthesis companies, so it is important that the Government supports small businesses in their efforts to comply with the screening Guidance. It would be helpful if the Government made available to us curated databases of prohibited sequences and provided an API for identifying matches to such sequences within a given sequence.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 1-10

Revised guidance to providers of synthetic DNA would ideally clear up some of the confusions inherent in the current list-based approach. The relationship between nucleic acid sequences and pathogens and toxins included on these lists is opaque. Some genes vary between listed and unlisted viruses in only a few point mutations. We can also see that there may not be a clear relationship between disease of concern and an agent; for example, the USA includes both *Bacillus anthracis* and *Bacillus cereus* biovar anthracis on the Select Agents list, but the CCL includes only *Bacillus anthracis*. Reducing confusion caused by the murky relationships between outcomes of concern (a disease, pathogen, or toxin is produced by a careless or malicious actor) and items on a taxonomically-based list will likely require a change in the focus of the guidance.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---



## 1-11

The current HHS Guidance focuses on identifying sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). Screening against FSAP- or CCL-listed organisms generally can be an effective approach to achieve biosecurity. However, recent advances in the technology making it more cost effective to synthesize gene length DNA sequences has made this approach incomplete, i.e. individual sequences not directly associated with an organism may in fact be a sequence of concern. To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

Contrary, we do not believe the following circumstances should require additional scrutiny:

1. A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.
2. A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 11:17:01 AM

Agency Type: Company/Business / Agency Other:

---

## 1-12

First, we would like to communicate our support for moving beyond a sequence screening approach that focuses on the Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). The current approach has become inefficient as the vast majority of sequences in an organism's genome do not 'endow or enhance' pathogenicity. Further, sequences of concern are being missed by those who are only applying the standards recommended in the 2010 guidance as there are sequence components outside of these lists that, alone or in combination, could be used for harm. We believe a more efficient and effective approach would be to focus on individual sequences as units of control rather than species and that this requires that the Department of Health and Human Services (HHS) to provide support by defining pathogenicity-linked sequences and providing contextual information describing the role of each sequence in a well-characterized pathogenic process. We further note that there is a lack of clarity around the definition of what constitutes a 'gene' for the purposes of export control. As part of your work to define how a synthetic genetic material provider can better identify sequences of concern, we urge HHS to ask the Department of Commerce to elaborate in updated Guidance as to what constitutes a 'gene' and to discuss the degree to which partial open reading frames or disrupted functional units can remove the requirement for an export license in some cases. As motivating examples, the following hypothetical orders might be said to fall under the definition of a 'gene' for the purposes of export control: A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein. A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence. A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct. In contrast, the following hypothetical orders may be considered to fall outside of the definition of a 'gene' for the purposes of export control: A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider can determine the preservation of functional elements. If this sequence is from a controlled bacterium, it is impossible to determine whether this sequence 'endows or enhances' pathogenicity. A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein. The customer provides literature citations to support an assertion that these changes disrupt the functional domain in a way that renders it no longer functional.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

---

## 1-13.1

I lead the R&D organization in a synthetic biology company and hold a PhD in this field, along with nearly 20 years of experience in this field. Our researchers work to bring quality and reliability in the field of synthetic biology so that our clients can pioneer innovative solutions. I appreciate the opportunity to comment.

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

On the other hand, we do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 11:53:58 AM

Agency Type: Company/Business / Agency Other:

---

## 1-13.2

To support the US bioeconomy this question is important for the synthetic biology field. I appreciate the opportunity to provide comment based on my 15 years in this field since completing my PhD and doing extensive research in this area. Working with customers firsthand provides unique insight into the importance of this question, and my hope that this comment will be enacted in final rule.

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

We do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 11:56:10 AM

Agency Type: Company/Business / Agency Other:

---

## 1-13.3

I represent clients in academic research and start up companies in this space working to enhance the US bioeconomy. We believe strongly that this is an important question, and all of the institutions appreciate the opportunity to comment. We need this to be enacted in final rule to level the playing field and make business predictable while supporting US bioeconomy growth. I personally have a PhD in this field and much of my research was in the area of synthetic biology.

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

We do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 11:58:41 AM

Agency Type: NGO / Agency Other:

---

## 1-13.4

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

We do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 11:59:16 AM

Agency Type: Company/Business / Agency Other:

---

## 1-13.5

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

Contrary, we do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 11:59:43 AM

Agency Type: Academia / Agency Other:

---

## 1-13.6

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

Contrary, we do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 12:00:12 PM

Agency Type: Company/Business / Agency Other:

---



## 1-13.7

In academia this question is critical. Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

Contrary, we do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 12:00:51 PM

Agency Type: Academia / Agency Other:

---

## 1-13.8

I have been working at companies in this area since the inception of synthetic biology. This question is very important for us to have answered in a definitive way, I appreciate the opportunity to comment.

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

Contrary, we do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 12:04:07 PM

Agency Type: Other / Agency Other: Consultant

---

## 1-13.9

Recent advances in synthetic biology technology has made it cost effective to synthesize gene length DNA sequences. For this reason, individual sequences not directly associated with an organism may in fact be a sequence of concern and hence, the Guidance should focus on more than just sequences derived from or encoding genes specific to Select Agents and Toxins listed by the Federal Select Agent Program (FSAP), or for international customers, items on the Commerce Control List (CCL). To better reflect the current state of technology and strengthen the biosecurity protections offered by the Guidance, we encourage HHS to not only focus on the agents and toxins on those lists, but to also focus on the individual sequences that may be pathogenic themselves.

To provide clarity as to when a sequence's pathogenicity is of concern, we recommend that HHS update the Guidance to include a definition of a "gene" of concern and require additional screening in the following circumstances:

A customer orders 75% of a controlled open reading frame, preserving important functional elements in the sequence, but not the entire encoded protein.

A customer orders a sequence that is a 'best match' to a gene from a regulated species but does not have what might be considered 'high' homology (e.g. 60% homology) to the controlled protein sequence.

A customer orders a sequence from a gene that would be subject to a license requirement but breaks these sequences up into short (e.g. 300 base pair) segments with homologous overlapping sequences, and indicates to the synthesis provider that they intend to assemble these sequences into a longer construct.

Contrary, we do not believe the following circumstances should require additional scrutiny:

A customer orders 75% of an open reading frame but due to a lack of annotation, neither the customer nor the synthesis provider is able to determine the preservation of functional elements.

A customer orders a full-length sequence but significantly changes a region that makes up a key functional domain in the protein, likely rendering the expressed product non-functional.

Submitted on: 10/25/2020 12:04:45 PM

Agency Type: Other / Agency Other: Association

---

## 1-19

Guidance should extend beyond current lists, however this should be done judiciously (described in next question). Several genes are also involved in the metabolic synthesis of controlled substances (for example opioid compounds). Genes might also come from regulated organisms and plants. Genes within these organisms might also be considered for regulation; however this might be redundant considering other enforcement mechanisms from agencies including the DEA. Moreover, pathways are multi-step processes, making it challenging to determine at which step to enforce screening.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 2-1

**Scope of the Guidance - Question 2: Are there potential benefits and/or downsides to screening for sequences not on the Select Agents and Toxins or CCL?**

---

Screening needs to be targeted to the threat. The threat is not just with the sequence but with the people handling the sequence. When there is an indication that a person is a threat, it may be beneficial to screen for sequences outside of FSAP list.

Submitted on: 9/15/2020 9:16:26 AM

Agency Type: Federal Government / Agency Other:

---

## 2-2

Yes. The organisms and toxins that are currently on controlled lists are there due to the unique characteristics that they possess that can enable effective weaponization. Thus, it is logical in the absence of man's ability to effectively perform genetic engineering to only regulate such organisms and toxins. However, we are now entering an era in which man has the ability to effectively transfer and express genes across the tree of life. While most transfers of genes between organisms are likely benign, there have been several documented cases in which the transfer of genes between organisms can result in gain of function. Manipulation of biothreat-encoding sequences (e.g., recombinant protein production, genome insertion, gene mutation, etc.), even for legitimate purposes, could lead to the production of novel or enhanced hazardous products. In fact, precedent has shown that genetic manipulation can lead to biodesigns with high pathogenicity, host bioregulation ability, vaccine escape capability, high transmissibility, high toxicity, controlled drug production capability, and species extinction capability.

While viewing this problem through an organism-level lens seems intractable, when viewed from the perspective of biothreat function the problem becomes substantially more palatable. More specifically, organism characteristics are the collective behavior resulting from the expression of their genomic material and hence the functions encoded in their genomes. By controlling sequences associated with biothreat functions (SoCs), or functions that result in adverse pathogen characteristics, one can reduce the biological risks associated with acquiring synthetic dsDNA, such as: (1) accidental/intentional gain of function, (2) acquisition of full genomes of controlled organisms or toxin sequences, and (3) the concern of actors by-passing the current Guidance (e.g., an actor ordering biothreat sequences from non-controlled organisms whose functions are identical to those from biothreats associated with controlled organisms). For example, the hemolysin E protein (an SoC associated with the damage biothreat function) is found in both *Escherichia coli* K-12 (a non-controlled strain) as well as *E. coli* O111 (a CCL-controlled strain) with identical activity.

There are numerous challenges associated with overhauling the fragmented curation practices and screening tools currently being implemented by dsDNA providers. Further, the necessary infrastructure build out to perform effective operational biosecurity could result in an inequitable situation for providers relative to non-screening vendors. More specifically, the self-imposed costs by vendors that developed an effective screening system based on the revised Guidance would have to be passed on to the DNA vendor (and presumably, customer). In addition, given the inadequate state of the current infrastructure, it would take an immense amount of time to develop the databases required for screening. These challenges may appear to be insurmountable, however, we propose a solution to this problem in our response to the final question.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 2-3

Yes, see previous response. One additional downside may be that whereas there are established legal/regulatory reasons for screening for select agents/toxins and the CCL (e.g. export control and SAR), there may not be for these additional sequences. It may be more challenging to justify and/or consistently/objectively evaluate the in-context risk of these additional prospective sequences. Most of the agents on the CCL and BSA are identified only by organism name, and yet the current guidance recommends that “housekeeping genes”, those not involved in the pathogenicity of the organism, be ignored. Making this determination is not always trivial, and is the main source of the added cost of follow up screening. Additional sequences of concern to be screened should be explicitly defined, along with homology criteria in order to simplify the process.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other

---

## 2-4

Some threats are difficult to include on any list due to the current insufficient level of understanding of the nature of the threat. Likewise, high-capability actors may be able to engineer sequences that would pass screening but still create a threat. For example, an actor might engineer an entirely artificial threat or change an organism's codon table to cause an unpredictable interpretation of a sequence. Function-based screening might possibly detect such threats, but is currently not mature enough to do so reliably.

Moreover, the number of actors capable of such actions is currently extremely small, however, and likely to stay that way for some time to come. Thus, it is currently better to focus on list-based screening with an expanded list (as described above).

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---



## 2-5

All viruses on the Select Agents and Toxins list and CCL should be screened. However, several additional viruses might still be judged risky enough to warrant being added to the DNA screening list, yet not meet the stringent criteria for addition to the Select Agents list, given the extensive requirements that it imposes on the approximately 250 registered facilities. Again, an organization such as the Global Virus Network could help HHS identify viruses that might fall into this category. The primary downside to adding additional sequences is that it does increase the size of the list and thus the possibility of additional yellow flags. However, if expansion of the list is limited to a short list of carefully selected viruses or newly identified toxins, the increase would be modest.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

## 2-6

Given modern synthetic biology's abilities to optimize phenotypic behavior (for good or bad) it is important to broaden our view of biorisk beyond the classical lists. Many non-listed organisms have the potential be tweaked to produce significant harm to humans. We should also widen our scope to include pathogenic risks to plants and animals of economic importance as well as materials degradation. Databases must be well curated to decrease false positives.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 2-7

When standardized criteria designed to illuminate the "dangerousness" of sequences are uniformly applied to all disease-causing organisms, it is readily observed that there are far more sequences of concern (SoCs) not on the HHS lists than from those on the lists. There is a limited number of organisms on the Select Agents and Toxins list and CCL, and a correspondingly small number of SoCs. To date on the IARPA Fun GCAT program, our team has annotated 278 SoCs from 16 species of bacteria on the HHS lists, and >1,600 SoCs from >70 species of bacteria not on the HHS lists. We have annotated 119 SoCs from most viruses on the HHS lists, and 144 SoCs from 29 viruses not on the HHS lists. We have also annotated all of the toxins on the HHS lists, as well as SoCs from 21 other eukaryotic species of pathogens of both plants and mammals not on the HHS lists. Further, given synthetic biology's abilities to insert, modify, or optimize functions within a wide variety of organisms, it is more important now than ever to broaden our view of biorisk beyond narrow lists containing a small number of known pathogens.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 2-8

Benefits: catch potentially harmful orders that are not explicitly included in the SA&T and CCL lists

Downsides: additional screening burden on synthesis companies. The Government should support the screening efforts of small companies

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 2-9

There are less clear regulatory justifications for screening for sequences unrelated to the Select Agents and Toxins List or CCL. However, screening for sequences not directly on these lists would allow capturing functionally similar but taxonomically distinct sequences of concern.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 2-10

Potential benefits of screening sequences not on current lists include the ability to capture any potential hazardous sequences pre-emptively. Moreover, sequences required for synthesis of controlled substances, including drug compounds (DEA enforced/regulated) or natural products likely do not fall under these lists. However, the downside of screening for enzymes within metabolic pathways is the need to define at which step of a synthetic pathway genes will be screened. Screening for genes upstream in many metabolic pathways will be highly prohibitive for engineers and/or biologists. Too downstream might result in insufficient screening and/or regulation.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 3-1

**Scope of the Guidance - Question 3: Should the scope of the Guidance be broadened beyond synthetic dsDNA? If so, how? Should the scope of the Guidance be broadened to other synthetic nucleic acids? If so, what synthetic sequences? Or, should the scope of the Guidance be broadened beyond providers of synthetic dsDNA? If so, to whom? Why?**

---

We suggest the sequence screening component of the Guidance should be focused on vendors of biological raw materials (i.e., DNA and RNA, including both single stranded and double stranded), but not technology providers that are up or downstream of the raw materials (e.g., raw material end users, biotechnology integrators, etc.). However, shorter oligonucleotides (

In contrast to the sequence screening component, we suggest that the customer screening component of the Guidance should apply to all synthetic biology companies, as customer verification is a useful biorisk mitigation measure. At the very least, customer screening should be performed by dsDNA providers because they provide one of the more refined basic, raw resources required for synthetic biology.

In addition, we also encourage US Government funded activities to require ordering of these raw materials from reputable vendors that follow the Guidance, which would further enable responsible research.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 3-2

Regarding other synthetic nucleic acids: There have been discussions around the screening of ssDNA (e.g. DNA oligos), ss/dsRNA, and/or other forms of ss/ds nucleic acids or their analogs. These are also worth considering. However, as stated above and below, adding these additional nucleic acid forms to the guidance should only be done post a deliberate/diligent cost/benefit analysis. Regarding beyond providers of synthetic dsDNA: Should the guidance be extended to other synthetic nucleic acids (see above), then presumably the scope of the guidance should extend to the providers of these other synthetic nucleic acids as well. There are other unit operations in the engineering biology cycle (beyond those involving the fabrication of synthetic nucleic acids) where similar screening approaches are feasible and could bring additional benefits. For example, DNA/RNA/protein sequence design/visualization and related information management systems (e.g. software platform infrastructure) could be given guidance. However, as stated above for broadening the scope of the screening beyond select agents/toxins/CCL, the cost/benefit analysis of this scope creep would need to be done in a diligent and deliberate manner. As has already been discussed in the context of synthetic dsDNA providers, it is also crucial to evaluate if the additional scope to the guidance would disproportionately be a burden to certain companies/entities. While it is only guidance (and not legislation/regulation), (U.S.) commercial software platforms may be compelled (for marketing/public relations reasons) to follow the best practices provided in the guidance, whereas other software tools (especially from the open-source/academic community) are unlikely to participate. These arguments/considerations are very similar to those discussed for dsDNA service providers (e.g. companies that screen vs. those that do not; benchtop oligo synthesizers that do not “phone home”).

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other

---



### 3-3

In a study we conducted in 2007, we estimated that 25 million short oligonucleotide sequences (oligos) were ordered each year, compared to only 50,000 double stranded DNA orders. In the intervening years, however, rapid advancement and innovation have changed the industry dramatically. Yet, while innovations in dsDNA synthesis have greatly changed, some facts about oligos remain the same. Short oligos hold very little information, so spurious hits will be common, increasing the cost and efforts to screen them. Our interviews with stakeholders suggested oligo synthesis is a low-margin and rapid delivery industry, and manual follow-up on flagged orders would unacceptably raise the price and time of delivery—calling into question whether oligo companies would follow any new guidance suggesting they screen orders. Although only longer oligos could be included (and thus reducing the number of false positives), we suspect the oligo synthesis industry would still resist screening oligos of this length. Even if oligos were included, adversaries could order oligos just below the guidance limit and ligate them, without being detected. Lastly, the inclusion of oligos does not cover PCR amplification. To capture efforts for the de novo synthesis of a virus via the assembly of oligos (as has been demonstrated), the system COULD screen oligo orders that involve many sequences that are each at least 25nt long OR many more shorter oligos (maybe a total order size of 1,000 nt), with thresholds set such that splitting orders between providers to avoid detection would be cumbersome and undesirable. Analyzing this type of order would provide enough information on the intent of the user, have overall a better margin than a single order, and be relatively rare (we think) so would not substantially increase the burden. Moreover, since the total number of oligos needed to assemble even short genomes is large, it would be impractical for a malicious actor to "split" the order amongst many providers. The main drawback of this type of system is that it would require producers who aren't currently covered by the guidance to implement it for relatively rare orders, so compliance may be an issue. Before this aspect were included in an enhanced guidance, a study should be performed to understand the current flow of orders through companies that make oligos and compared to the materials needed for the synthesis or engineering of Select Agents to set the thresholds for screening. Overall, and for the same reasons that drove us to our conclusion in 2007, we believe individual oligos are unsuited for screening, and efforts to screen DNA products should focus on gene synthesis and other larger products. We do believe it is feasible and perhaps desirable to screen a set of oligo orders from a single customer to prevent the synthesis of pathogens de novo via that route.

Submitted on: 10/23/2020 1:19:36 PM

Agency Type: Company/Business / Agency Other:

---

## 3-3

Improved assembly protocols now allow large-scale genetic constructs to be readily assembled from short oligos (40 – 60 bp). As such, ssDNA poses just as much of a threat as dsDNA and should be similarly controlled. Signature-based detection methods (see below) now allow effective screening of short sequences, so screening ssDNA can be practical as well.

Synthetic RNA is not currently as widely used, but can potentially be transformed into DNA via reverse transcription. While this is not a typical current practice to create synthetic DNA, it offers a relatively accessible route to evading DNA screening. Thus, the guidance should be applied to synthetic RNA as well.

Non-standard nucleic acids (XNA) are likely not yet common or accessible enough to pose a significant threat. Progress in this field should be monitored, however, as another possible workaround. Of particular import would be the emergence of significant commercial XNA vendors, since that will greatly increase the number of potential actors capable of working with XNA.

Finally, the guidance should be applied not only to providers of synthetic DNA but also to those providing services that produce larger-scale synthetic constructs and organisms. Notably, this should include companies that provide gene editing services, manufacturers of desktop synthesis or editing equipment, organism engineering companies, and cloud laboratories.

Note that this might result in a sequence being screened multiple times during its production (e.g., at a synthesis company who makes oligos for an editing company that in turn uses them to produce a strain for one of its customers, then again by a cloud laboratory where the customer sends the resulting organism for experiments). Multiple screening is desirable because it provides defense in depth. At the same time, none of the companies involved can safely assume that the screening is completed elsewhere (e.g., the editing company cannot assume safety of a transformed sequence without checking it because the synthesis company does not know the sequence context of the transformation, and the cloud laboratory cannot assume that it has not been sent a dangerous sample with an incorrect label).

Screening genome-scale sequences would pose an undue burden with BLAST, but emerging methods from the IARPA FunGCAT program are capable of screening several orders of magnitude faster without an increase in false negatives. This allows the screening of genome-scale sequence data in only a few minutes with moderate computational resources. These methods can be applied to both assembled sequences and raw sequencing data. Adoption of methods such as these would allow the guidance to be applied to strain designs or samples without posing an undue burden.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 3-4

The Guidance should pertain to other types of synthetic DNA and RNA and to recommend sequence screening of oligonucleotide pools.

As acknowledged in the RFI, the Guidance currently addresses only synthetic double-stranded DNA. The conversion between single- and double-stranded DNA is straightforward and carried out frequently by synthesis providers. RNA can also be converted back and forth to DNA using off-the-shelf reagent kits. Given this, the Guidance should be broadened to cover manufacture of synthetic DNA in general, regardless of strandedness, along with synthetic RNA.

Additionally, the scope of the guidance should be expanded to cover recommendations around sequence screening for pools of shorter oligonucleotide sequences. The current Guidance draws a line at (a somewhat arbitrary) 200 base pairs, only asking synthesis providers to evaluate risk for sequences of that length or longer. DNA synthesis providers, however, routinely use shorter oligos (of 50-100 base pairs) to assemble and manufacture gene-length DNA sequences. Given the ease with which oligonucleotide pools can be used to produce these longer sequences, it is important that the Guidance both a) recommend the screening of these pools of shorter oligonucleotides while b) explaining that the 'best match' approach is not appropriate for individual, shorter DNA sequences (as this approach results in a high false positive hit rate). Instead, the Guidance should recommend the use of de novo sequence assembly strategies (derived from next generation sequencing analysis approaches) as one way to estimate whether a pool of oligonucleotides could be used to assemble a gene-length fragment. If such an approach does detect a potential contiguous assembly (a 'contig'), the Guidance should further recommend that these contigs be subject to 'best match' sequence screening.

Many DNA synthesis providers also manufacture and sell proteins derived from synthetic DNA. Under current Guidance, it may not be clear to the provider whether regulatory control of DNA implies any such control of expressed protein (outside of controlled toxins, which are self-evidently capable of causing harm). The Guidance should be expanded to outline that expressed protein does not represent the same degree of risk of misuse given that it cannot be easily replicated without access to the encoding DNA or RNA. Providers should still determine whether a requested protein poses any biosafety risk.

In addition to a focus on the threat of misuse, the Agency should also consider extending the scope of the Guidance to also provide recommendations to DNA synthesis providers on how to approach the synthesis of genes involved in the pathways that produce small molecules subject to regulatory control, i.e. schedule 1 drugs under the U.S. Controlled Substances Act. This will aid in preventing the illegal manufacturing of illicit substances. Given that, at present, these genes are not subject to regulatory control, NAME recommends that the U.S. government clarify that carrying out customer screening as directed by the Guidance provides an important risk reduction in accepting and fulfilling these orders.

Broaden the scope of the customer screening portion of the guidance to apply uniformly to all synthetic biology companies

In terms of broadening the scope of the Guidance beyond providers of synthetic DNA, the customer screening and records retention portion of the Guidance should be universally recommended as a best practice for synthetic biology companies more broadly. Regardless of the step in the value chain carried

out by a given company, it is still important to 'know your customer' - the Guidance should make this recommendation explicit.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 3-5

We believe that the Guidance needs to be expanded to include providers of synthetic oligonucleotides. Protocols for constructing viral genomes, toxin genes, or bacterial genes from synthetic oligonucleotides larger than 40 to 50 bases are now available and clearly explained in the scientific literature. Undergraduate students are easily capable of assembling a poliovirus genome from a few hundred oligonucleotides using a small number of reaction tubes. As discussed in greater detail in an answer below, the computational technology now exists to reliably screen oligonucleotides as short as 50 bases. Screening of oligos between 40 and 50 bases is still possible, but with significantly higher false negative or false positive rates.

Advances in synthetic biology have made the assembly of DNA molecules the size of small and even large viral genomes from synthetic double-stranded DNA or DNA oligonucleotides simple for laboratory workers of ordinary technical skill. Similarly, viral reverse genetics reactions in which infectious virus is produced by installing viral genes and/or genomes in appropriate cells is relatively straightforward for any molecular biologist with access to cultured cells and freely available scientific literature. Once a worker has the needed synthetic DNA, the laboratory operations one would perform to produce virus probably would look just like ordinary lab activities going on in academic and industrial labs. Thus, virus or toxin production could go on in full view of others in a lab and not raise any alarms.

However, assembly of potentially infectious viral genomes requires high-quality double-stranded DNA (the sequence must be correct) or high-quality oligonucleotides, i.e., of the quality only the best commercial providers deliver and beyond that of a current typical bench-top synthesizer. Assembly of genes or genomes using anything other than very high-quality DNA will almost certainly produce genomes with sequence errors that would not be useful in the production of infectious virus.

We believe that synthetic RNA should be screened, as well. While RNA is more difficult to work with and more expensive to construct than DNA, an order for a complete viral genomic RNA should be detected. At present, it would be difficult and expensive to construct genes or genomes from RNA oligonucleotides, but technologies could evolve to make this simpler and more accessible.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

## 3-6

Yes, we believe the scope should be broadened from dsDNA synthesis to include other synthetic nucleic acids. For example, it should cover products of genome engineering products and associated tools. It ought to include the new abilities of the whole genome editing industry, as their new tools may be located at users' facilities rather than being clustered within large synthesis factories. The guidance should address biosecurity challenges associated with genome editing such as predicting functional impact edits in the context of the living organism, thus addressing altered protein interactions and regulatory elements. Furthermore, the guidance should address the need to assess potential biorisk of combinatorial edits in the context of a living organisms.

The scope should be broadened for applicability to providers of genome engineering products, tools, and services.

Rapid advances in benchtop dsDNA synthesis machines (and other synthetic nucleic acids) should also be explicitly addressed in the new Guidance.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 3-7

Yes, the scope should be broadened to extend to all synthetic nucleic acids and proteins, since biotreats are not limited to dsDNA. We believe a wholesale approach for identifying, categorizing/describing, and databasing sequences of concern based on objective criteria should be devised and widely shared among responsible companies and nations, so that it may be used to keep track of the use of these sequences in research and synthesis. Our opinion is that these sequences should not be banned from proper use, but rather monitored for misuse. Monitoring requires knowing what sequences might be misused.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 3-8

Absolutely-- the Guidance should cover synthesis of any information-carrying polymer that can be converted into dsDNA with good fidelity. In particular, no differentiation should be made between dsDNA and ssDNA, RNA, or XNA, etc. This is essential given the expected improvements in ssDNA and RNA synthesizers, where direct synthesis of gene-length ssDNA will likely be enabled over the next decade, and could easily be converted into dsDNA.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---



## 3-9

Since molecular biology techniques allow various synthetic nucleic acids to be transformed into one another, the screening guidance should be broadened to apply to other nucleic acids such as synthetic RNA and ssDNA. The guidance should also be broadened to cover oligo pools that might be assembled into sequences of concern.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 3-10

We greatly appreciate that you invite comment on the current focus of the guidance on only synthetic double-stranded DNA. Single- and double-stranded DNA conversions are commonly carried out and additionally, RNA can also be interconverted to DNA using off-the-shelf reagent kits. We also recommend that HHS extend guidance to include screening of oligonucleotide pools. Providers of synthetic genetic material regularly use oligos smaller than 200 nucleotides to assemble and manufacture gene-length DNA sequences. Given the ease of performing these tasks, we strongly urge HHS to broaden the guidance to include all types of synthetically generated DNA and RNA, along with pools of shorter oligonucleotide sequences.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

## 4-1

**Scope of the Guidance - Question 4: Should the scope of the Guidance be narrowed, either in terms of types of sequences screened or the audience of the Guidance? Why or why not?**

---

The Guidance needs to be based on risk and not technique. The guidance needs flexibility that can be built in by the biological and biosecurity risk assessment process.

Submitted on: 9/15/2020 9:16:26 AM

Agency Type: Federal Government / Agency Other:

---

## 4-2

As mentioned in our previous responses, the scope of the Guidance when considering types of sequences screened is tied to the Government's decision on the fundamental unit of control. Our suggested fundamental unit of control (biothreat function) would result in both an expansion and contraction of the scope for sequences screened. For example, the Guidance would be expanded to all potent toxins that impact humans such as sarafotoxins—a toxin lethal in mice on a per weight basis similar to ricin toxin—not just currently controlled toxins. In parallel the Guidance would be contracted, as there would be no need to review housekeeping genes from controlled organisms (e.g., DNA polymerase). The greatest inefficiency associated with the current Guidance is the incredible amount of ambiguity and inhomogeneity in the definitions and rules required for screening. If the sequence screening related Guidance were to be transformed into a "recipe book", then much of the scope-induced inefficiency issues would be mitigated.

Regarding the scope of the audience of the Guidance, see our response to the previous question.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 4-3

There could be arguments for narrowing the sequence screening in the sense of not screening against all sequences within an organism found on the SAR/CCL lists. For example, explicitly exempting in the revised Guidance (or associated materials/databases) the genes within each organism that are not of concern/interest (e.g. those not related pathogenicity, etc.). The current scope of the audience of the Guidance (providers of synthetic dsDNA) is already quite narrow.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other

---

## 4-4

Current guidelines call for a system that flags genes that are not associated with pathogenicity if they are a "best match" to a sequence from a Select Agent. Often, questions arise about whether this is a flaw in the guidance or a feature. In considering this issue, we note that we view the purpose of the existing guidance as twofold:

To prevent malicious actors from acquiring synthetic DNA that they could use to cause harm directly

To detect (potentially) malicious research activities before actors gain the capacity to do harm  
In our analysis, we consider the flagging of housekeeping genes to be a feature that supports the second aim of the guidance above, a conclusion based largely on discussions with the security community. The following example lays out a scenario using the screening system with housekeeping genes: Imagine that a previously unknown customer orders the full open reading frame (ORF) of DNA gyrase identical to that in *B. anthracis*. Under the guidance, screeners should:

Verify no other *Bacillus* species match better or as well. (i.e., know what you are making)

Conduct a customer screen to reveal more about the customer. (i.e., know your customer)  
With the information about the order and the customer, should a responsible company fill the order? Internal secondary screening could reveal more information about the customer, such as their identity as:

A hobbyist working in a DIY bio community

A new biotech company without a website

A graduate student working in a lab that studies yeast chromatin structure

A state-owned company operated by a US adversary  
From here, a responsible company should consider available context when filling the order. If a customer is ordering only this gene, or perhaps this gene and others from *anthracis*, then the cause for alarm may be elevated, as there is limited scientific purposes for studying the housekeeping genes of pathogens. In other words, flagging housekeeping genes as of concern may help detect malicious use. Follow-on analysis may reveal a legitimate scientific purpose for this order. For example, the customer may be ordering the DNA for every example of DNA gyrase known for Structure/Activity Relationship studies. Beyond this effort, however, what is the scientific purpose of studying the housekeeping genes of THIS pathogen, specifically? In contrast, there are many scientific reasons for studying the genes of a pathogen that confer pathogenicity. To study the mechanism of action of lethal factor or edema factor, or to develop specific countermeasures to it, a researcher really must work with those genes. In our experience, most companies that are reporting many spurious hits from housekeeping genes are not following the guidance as written. These companies have reported to us that they check the sequence and if a DNA sequence from a Select Agent is above a certain threshold, they consider it a hit. Some companies simply determine if a Select Agent sequence is on the "first page" of BLAST hits. Other companies use explicit thresholds for identity/similarity and if a Select Agent gene passes this threshold, it is a hit. In both cases, you would expect highly conserved housekeeping genes to be flagged. The guidance was written the way it was to eliminate hits in highly conserved housekeeping genes. The government should remind industry that the guidance is intended to be used a system and if they do not use a best match approach, spurious hits

from conserved housekeeping genes will be common. If they want to reduce flags from housekeeping genes, they should follow the entirety of the guidance.

Submitted on: 10/23/2020 1:19:36 PM

Agency Type: Company/Business / Agency Other:

---

## 4-5

Many of the individual genes of bacterial and eukaryotic pathogens are not of concern and are a common source of false positives. The current guidance exempts "house-keeping" genes without specificity, which poses a potential problem for screeners in determining what can be exempted while maintaining compliance. Many false positives, however, are on highly conserved genes with no significant relationship to pathogenicity (e.g., rRNA, tRNA, rpoB, dnaA). Specifically excluding even a relatively small number of such genes may have a significant impact on false positive rates. The classes of gene to exclude, however, will differ between classes of organism: for example, translational mechanisms are not typically of interest for eukaryotic pathogens, but highly significant for viral pathogens. We believe that the government should convene a standing committee of scientific experts in the relevant organisms to create and maintain a list of specific exclusions.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---



## 4-6

No, we feel it should not be narrowed. The current Guidance was suitable for 2010. The field of synthetic biology has advanced greatly since then and the new Guidance should not be solely for the gene synthesis providers. It should consider the advances in genomic engineering from individual gene to whole genome level, as well as benchtop genomic synthesis and modification and provide them with appropriate guidance.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 4-7

The scope of the Guidance should be broadened rather than narrowed. Something on the order of 2,000 microorganisms are capable of causing disease in humans, including approximately 600 fungi, ~600 bacteria, ~300 helminths, ~300 viruses, and >50 protozoa. We have curated sequences of concern in >120 species to date, and our work is not finished yet. Our annotation of the existing literature is nowhere near complete, and future research will greatly expand the number of sequences available for annotation. Even this number neglects the microbes that can cause disease in livestock and crops essential for human existence and economic life.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 4-8

The guidance should not be narrowed. However, it needs to be more precisely specified-- for example, how does the screener determine what is considered a “housekeeping gene” in any given infectious agent and what is not?

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 1-1

**Sequence Screening - Question 1: Should the Guidance be further clarified or otherwise updated to identify embedded “sequences of concern” within larger-length orders? If so, how?**

---

Yes. Referring back to our response to the first question, if the Guidance were to be revised to define the fundamental unit of control at the biothreat function-level, then all potential open reading frames should be analyzed for functional biothreats. There are several bioinformatic techniques that can be used to identify potential open reading frames. One such approach is to leverage the aligner results from protein databases to identify statistically relevant potential open reading frames. More specifically, one could group protein alignment results by frame and histogram the start and stop positions of alignments across the input sequence weighted by alignment percent identity. Clustering the weighted histogram frequencies, would allow for quantitatively defining the boundaries of potential open reading frames or regions per frame. Once the regions are defined, alignment data can be grouped and analyzed per region. Alternatively, open reading frame finders exist that could be used for this same purpose. Once all of the regions associated with the input sequences of a given order are defined, one can group the regions

that map to the same SoC and, using subject start/end data from each region, assemble potential embedded biothreat sequence(s). These potentially embedded sequences can subsequently be rescreened to determine if they pass the Guidance flagging criteria. It is worth noting that this analysis can be done for orders of any length.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Orgnaization

## 1-2

There is no need for further clarification. The 200-bp (66-aa) windowing approach is an effective means to identify embedded sequences. It is another question if the window length should be reduced; however, this would need to be assessed from a cost/benefit analysis perspective. Should the guidance be revised to suggest how multiple windows in the same sequence (or within a set of sequences) could impact the assessment (e.g. two interspersed sequences that are not independently best matches but when concatenated/reordered do become a best match), clarification with examples of how best to do this would be very useful.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-3

DNA editing tools now make it relatively simple to extract portions of a nucleic acid sequence, which may later be assembled to produce a threat sequence. Therefore, screening should consider embedded sequences as well as complete sequences.

The screening length needs to be reduced from 200 bp to a significantly shorter length, due to the widespread availability of methods for assembling large sequences from short fragments (e.g., oligos). Results from performers in the IARPA FunGCAT program suggest it is likely to be feasible to require screening for sequences as short as 50 bp. On the higher side, some commercial dsDNA vendors already routinely perform screening on all sequences of a least 75 bp. Thus, the minimum size should be somewhere in the range of 50-75 bp.

The guidance should also clarify that for "each ... nucleic acid segment", the segments should be significantly overlapping, such that an adversary cannot easily evade detection just by placing sequences of concern at likely boundary coordinates. Ideally, every possible segment should be checked, i.e., a sliding window; this would be burdensome for BLAST, but is readily possible with signature-based detection (see below).

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 1-4

It is possible to divide controlled sequences up into much smaller (~30-50bp) segments, space them out across a longer construct, and include restriction sites and other sequence motifs that would make reassembly of these small sequence portions from controlled genes relatively straightforward. Most alignment algorithms (like BLAST or k-mer based aligners like bwa) work via seed-and-extend strategies, by which the aligner finds all possible matches of some very small seed sequence (e.g. blastx defaults to a match of 3 amino acids [translated from 9 nucleotides in the query sequence]). Aligners can be parameterized with these small seed lengths to find all sequence regions in a query sequence that have a 'best match' over anything equal to- or longer than the seed sequence to a controlled organism. Given these search strategies, instead of asking 'is there a contiguous 200bp region with best match to a controlled organism?', screening systems can ask the more sensitive question: 'Adding up all of the sequence detected in this construct as a) longer than the seed length and b) unique to a specific controlled organism, is there any single organism with more than 200 possibly-noncontiguous base pairs of unique sequence in this construct?'

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-5

For the reasons discussed above, we believe that the embedded sequence screening window length should be shortened significantly below the current 200 bp, even to as low as 40 to 50 bp.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---



## 1-6

Modern screening technology is fully capable of detecting embedded “sequences of concern” within larger-length orders. See answer directly below.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 1-7

Detecting threatening functions within short sequences is an important consideration for analyzing larger-length orders. Under the Fun GCAT program, we have developed an open source bioinformatics workflow called SeqScreen (<https://gitlab.com/treangenlab/seqscreen>) to assign Functions of Sequences of Concern (FunSoCs) to short gene sequence fragments. Testing and evaluation of SeqScreen software has shown that sequences as short as 40 nucleotides can be accurately detected and characterized. We are currently evaluating the best protocol for subdividing larger-length sequences into smaller ones for optimal processing (e.g., shorter divisions of a fixed length, open reading frame predictions within long sequences).

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 1-8

200 bases is too long of a window; the Guidance should concern the Best Match over each 100 base segment. Some peptide toxins, whose insertion into an infectious agent could dramatically increase the danger associated with the agent, are ~30 amino acids (90 bases) long. Furthermore, it is possible (and currently is the norm) to assemble complete genes from 80-150 nt long oligonucleotides, so screening of such shorter sequences should obviously be included in the Guidance. Otherwise, there will be a giant loophole where the Guidance does not protect against malicious or accidental misuse of DNA synthesis equipment in cases where the user assembles a longer dsDNA from short oligonucleotides.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 1-9

One example of embedded sequences of concern might include known enzymatic cleavage tags for human proteins that are non-standard in biochemistry techniques. This might pose a concern whether the proteolytic processing of viral protein precursors are being manipulated. Of concern might be obscure enzymatic cleavage tags for enzymes within humans.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 2-1

### **Sequence Screening - Question 2: Are there approaches other than the Best Match, using the Basic Local Alignment Search Tool (BLAST) or other local sequence alignment tools, to check against the National Institutes of Health's (NIH's) GenBank database that should be considered? What are the benefits and/or downsides of those approaches compared with the current Guidance?**

---

Referring back to our response to the first question, if the Guidance were to be revised to define the fundamental unit of control at the biothreat function level, then a new criterion will be required. However, this criterion should be broken into two parts: (1) it should clearly define the database sources to be used and (2) it should quantitatively define the flagging requirement for review. Regarding the former part of the criterion, we suggest leveraging the UniProt Knowledgebase (UniProtKB) since it is the leading data source for functional information on proteins, and it defines functional units of coding sequences such as chains and active peptides (<https://www.uniprot.org/help/uniprotkb>). Further, UniProt removes large redundancies found in other protein databases such as NCBI's nr database through careful annotation and sequencing clustering (e.g., UniRef100). Since UniProt only enables screening for coding sequences, NCBI's nt database, which includes GenBank's database, can be leveraged as well to provide better granularity for taxonomy, as such information is relevant to identify contextual SoCs as described above (to identify regulated pathogens). However, due to the large size of nt and redundancy with UniProt, the nt database should be filtered to remove RNA information (RefSeq, Protein Databank, Protein Information Resource, and Protein Research Foundation) as well as non-useful nucleotide sequences for dsDNA screening (Third Party Annotations, Synthetic Sequences, Patents, etc.). Additionally, these databases should be "cleaned" to remove ambiguous sequences, such as those containing undefined nucleotide and amino acid sequences (N's and X's, respectively), which will reduce false positives during sequence screening.

It is important to note that we have developed a biocuration pipeline for cataloging, updating, and maintaining all of our databases required for sequence screening, including the continuous expansion and quality assurance of our SoC database. Our curation pipeline is accompanied by a Standard Operating Procedure to provide standardization across curator entry. Entries are compiled within an access-controlled database and identified from publicly available information sources. A SoC is only included in the database if its sequence encodes for a verified function based on experimental data from the literature or (in cases such as some select agent viruses where experimental data do not exist) based on homology to a sequence encoding a verified function. Protein sequences are retrieved from UniProt when available or manually entered based on literature documentation. Functional metadata categories described in our response to the first question were developed based on panel discussions of high-level biothreat functions used by pathogens and organisms producing toxins, drugs, and bioregulators. Further details can be provided upon request.

As for a quantitative definition for the flagging requirement, we suggest criteria similar to the following:

(Note: the values presented in items 4 and 7 are shown as an example of a quantitative Guidance. The actual values should be set based on an analysis of historical customer order data in which metrics, such as false positives and false negatives, are balanced to provide the greatest risk reduction with minimal burden to the vendors.)

Identify all regions (e.g., open reading frames and untranslated regions) associated with the input biological sequence (see response to previous question for more details);

Group alignment results by frame and region;

For each region score the associated alignments based on the multiplication of the alignments percent identity and percent coverage of the region;

Define a hit to a SoC to be when an alignment associated with a SoC has a score within 10% of the top scoring alignment in that region;

Loop through all regions to determine if any region is associated with a hit to a SoC. If so, first check if any of the SoC hits are associated Tier 1 (as defined in our response to the first question), if so, trigger the sequence for review;

Else if the biothreat function is associated with Tier 2 (as defined in our response to the first question), then check if the sequence has greater than one region. If so, check if any of the non-contextual biothreat regions is associated with a lowest common ancestor taxid of a regulated pathogen – the set of taxids used for the lowest common ancestor calculation should be those associated with alignments that have scores equal to the top score;

Else check customer history for the percent of SoCs ordered from the same pathogen (biothreat function associated with Tier 3, as defined in our response to the first question). If the customer has ordered greater than 25% of the SoCs from the same pathogen, then trigger for review and increase the historical trigger threshold by 25% - note historical data should be restricted to eight years' worth of customer orders.

The benefits to the suggested rules-based flagging criteria is that the data sources are clearly defined, and the flagging criteria is clearly presented as a decision tree that is defined quantitatively, enabling

direct implementation by experts. The downside to the criteria is that it requires an overhaul of the biosecurity infrastructure currently being implemented by the industry.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 2-2

The best match approach is a very sensible one, in the absence of a more complete understanding of which genes are of concern or which sets of genes in combination (as an emergent property) are of concern. However, with a more complete understanding, the best match approach could lose its effectiveness. In addition to the best match approach, were there to be a curated database of sequences of known interest and/or combinations of genes of known interest, perhaps these curated databases could be screened first (with a thresholding approach to alert the screener above a certain percentage sequence identity or fraction of gene sets present) to assess where a more complete understanding is available, and then fall back on the best match approach where there is less complete understanding. The downside to best match is that it becomes increasingly computationally expensive as the size of the genetic databases screened against increase, and that it can be weak/defeated with some effort if there are gene orthologs in the database that have not been noted as having been of concern. The upside of the best match approach is that it can help reduce downstream false positive follow up costs, and can provide a helpful signal in the absence of a more complete biological understanding. The curated database(s) approach is advantageous in that it can use our more complete understanding of (emergent) biological function when available, and that screening these relatively small curated databases is much faster and less expensive than screening (via best match approach) the entire contents of public gene databases. The downside of the curated database(s) approach is that someone or some entity (private or public) will need to create, maintain, and continually improve these databases, and that they may pose as a information hazard since all sequences of interest (or combinations thereof of interest) are all together in one place. Until our biological understanding is more complete, perhaps combinations of best match and curated database approaches are well advised, supplementing with predictive bioinformatics tools (see below) where available.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---



## 2-3

An important alternative to sequence alignment is signature-based detection. Signature-based detection checks for exact matches to a template "signature" rather than approximate matching to a region. This signature may be a complete k-mer or it may include variable or unknown regions and is typically significantly much shorter than regions searched for by BLAST-based methods. The important difference is that a signature exactly specifies which portions of a sequence are expected to be important for making judgements and which portions are not. This can allow greatly reduced false positive rates and greatly reduced computational cost in evaluating a sequence.

Signatures are already used in the HHS select agents and toxins list. The definition of conotoxins to control specifies a particular amino acid sequence ("X1CCX2PACGX3X4X5X6CX7"), in which the "X" values are unknown amino acids. Signature-based detection can use this template directly, matching precisely those organisms that are controlled. BLAST, however, might well match when a sequence begins with "AQG", even though this cannot be a controlled conotoxin. At the same time, BLAST could be decoyed into mis-categorizing a sequence as non-controlled based on matches in the "X" values.

Signatures have not previously been used much in biological safety screening because developing them by hand has required significant investments of time and expertise. Signatures are widely used in cyber-defense, however, as even highly variable self-concealing malware tends to have critical conserved elements that can be readily identified with the aid of automation.

Performers in the IARPA FunGCAT program have demonstrated that the same applies to biological threats. Conserved sequences associated with pathogenicity can be used for highly effective screening: less than 1% false positives on sequences 50 base pairs or more. This is at least a three-fold improvement over false positive rates from BLAST, with no increase in false negatives, despite moving to a shorter window. Developing these signatures can be cost-effective with the aid of automation, and a software product providing signatures to screen for all controlled elements on the BSAT list is being made available commercially.

We thus believe that guidance be adjusted in the following ways:

The government should not specify a particular technology to be used for determining whether a sequence is controlled. Instead, screening systems should be evaluated in terms of their performance against a "gold standard" test set maintained by a neutral third-party governmental organization such as NIST.

Minimum sequence screening length should be shortened significantly below 200 bp, perhaps even to as low as 50 bp.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 2-4

The 2010 Guidance does not mention what has become a key data source for sequence screening: the [National Center for Biotechnology Information \(NCBI\) Taxonomy Database](#). GenBank records are each linked to a species of origin using a taxonomy ID, essentially a location in the tree of life maintained by NCBI. When a 'hit' comes back from a Best Match-compliant screening approach, the taxon ID in the GenBank record is evaluated against a curated list of taxon IDs that align with named species on the various control lists. The taxon ID, then, is the key link between sequence and organism-based lists. The Guidance should consider going a step further and providing NCBI taxon IDs for each of the species on the FSAP and CCL control lists, to eliminate the need for providers to perform this mapping (and perhaps introduce mistakes in terms of the level of granularity intended by regulatory authorities). For example, the entry for *Escherichia coli*[1] contains 3,330 strains. The CCL indicates that Shiga toxin-producing strains of *E. coli* (STEC) in serogroups O26, O45, O103, O104, O111, O121, O145 and O157 and 'other' shiga toxin-producing serogroups are controlled. It is left up to the DNA synthesis provider to map the CCL language to some number of the 3,330 strains in the NCBI Taxonomy Database and the more than 3.8 million individual unique protein records from *E. coli* strains in RefSeq.

Using the whole of Genbank for alignment, as suggested by the 2010 Guidance, results in a large number of alignments to identical sequences - wasting computational resources and requiring additional interpretation to determine and discard identical alignments. Thus, we further recommend the Guidance be more specific about the data source for sequence alignment and suggest use of the NCBI 'nr' and 'nt' databases. These databases are sets of all known non-redundant protein and nucleotide sequences (respectively), each of which is annotated with a species of origin using NCBI taxonomy ID numbers. The non-redundant nature of these datasets maximizes the diversity of protein or nucleotide sequences searched (to ensure a detected sequence of concern is truly unique to a controlled pathogen) while minimizing the total number of sequences against which a 'best match' approach will need to align.

The Guidance should further describe why alignment to a comprehensive database like nr is necessary, compared to an alternative approach of aligning to just a database of 'bad' sequences. If alignment is carried out only against a list of controlled sequences, it is impossible to determine whether a customer's sequence is more similar to a controlled sequence than to a similar sequence from a non-controlled organism. It is also impossible to determine whether that sequence happens to not be unique to the controlled organism in question. Given that regulatory control and license requirements are only triggered when sequences are unique to controlled pathogens, the Guidance should be clear that sequence screening provides unambiguous results only when a comprehensive sequence database like nr is used.

In addition to updating the Guidance, HHS should invest in improvements to nr and nt by creating a secondary database of preprocessed sequence regions that are known to be unique to controlled organisms across very similar families of proteins and functional RNA sequences. This metadata would

make provider interpretation of customer screening results more accurate since an important consideration for regulatory control is whether the particular region of sequence ordered is unique to a controlled organism. Using currently-available public resources this is often a complicated question to answer.

The 'best match' approach remains a reasonable high-level description that is independent of the specific algorithm used. This is beneficial and has permitted innovation in methodologies since 2010 while preserving the quality of analysis findings. The government should consider extending the definition of 'best match' to include a discussion of the idea behind a 'culling limit'. This is a parameter in the BLAST family of alignment algorithms (with cognates in other alignment algorithms) that limits the number of alignments returned that share start and stop coordinates (or that are entirely contained within other alignments). This is a crucial parameter when searching large databases to ensure that common sequence motifs do not end up returning thousands of alignment results, drowning out shorter hits to less common sequence that may be a 'best match' to sequences of concern.

[1] <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=562>

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 2-5

An important alternative to BLAST (and similar local sequence alignment techniques) is signature-based detection. Signature-based detection checks for exact matches to a template "signature" rather than approximate matching to a region. This signature may be a complete k-mer or it may include variable or unknown regions. The important difference is that a signature exactly specifies which portions of a sequence are expected to be important for making judgements and which portions are not. This can allow greatly reduced false positive rates and greatly reduced computational cost in evaluating a sequence.

Signatures are already used in the HHS select agents and toxins list. The definition of conotoxins to control specifies a particular amino acid sequence ("X1CCX2PACGX3X4X5X6CX7"), in which the "X" values are unknown amino acids. Signature-based detection can use this template directly, matching precisely those organisms that are controlled. BLAST, however, might well match when a sequence begins with "AQG", even though this cannot be a controlled conotoxin. At the same time, BLAST could be decoyed into mis-categorizing a sequence as non-controlled based on matches in the "X" values.

Signatures have not previously been used much in biological safety screening because developing them by hand has required significant investments of time and expertise. Signatures are widely used in cyber-defense, however, as even highly variable self-concealing malware tends to have critical conserved elements that can be readily identified with the aid of automation.

Performers in the IARPA FunGCAT program have demonstrated that the same applies to biological threats. Conserved sequences associated with pathogenicity can be used for highly effective screening: less than 1% false positives on sequences 50 bp or more. This is at least a three-fold improvement over false positive rates from BLAST, with no increase in false negatives, despite moving to a shorter window. Developing these signatures can be cost-effective with the aid of automation, and a software product providing signatures to screen for all controlled elements on the BSAT list is being made available commercially.

We recommend two changes to the "Sequence Screening" section of the Guidance:

The Guidance currently states: "The U.S. Government recommends that providers select a sequence screening software tool that utilizes a local sequence alignment technique; a popular and publicly available suite of algorithms that meets this requirement is the BLAST family of tools, and other tools are available". Such a recommendation discourages the development and deployment of potentially superior approaches, such as signature-based detection methods described above. The Guidance should no longer specify a preferred technology, rather should be performance based, ideally with specific criteria included, such as false-positive and false-negative performance. In an answer to a

question below, we describe how NIST might be tasked to assist HHS in performance evaluation, for example, by providing test sets for evaluation or by testing screening software itself.

We believe that the embedded sequence screening window length should be shortened significantly below the current 200 bp, even to as low as 40 to 50 bp.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

## 2-6

[SeqScreen](#) is an open-source software platform that connects a number of open source tools and databases for the purpose of predicting the taxonomy and function of a query sequence.

<https://gitlab.com/treangenlab/seqscreen/-/wikis/01.-SeqScreen-Overview>

Results include pathogenicity predictions incorporating machine learning functional algorithms along with match-based sequence screening.

This software can aid detection of biorisk in novel sequences. This resource could be improved to meet the screening needs of the genome engineering community by including prediction of the functional impact of edits in a genomic context, inclusion of shorter sequences, end user decision criteria, etc. We encourage Government to support ongoing maintenance and improvement of this resource, coordinating with industry end-users.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 2-7

The SeqScreen workflow utilizes BLASTX and BLASTN in “sensitive” mode and DIAMOND in “fast” mode. In fast mode, DIAMOND results are parsed by considering only those matches that are within 1% of the highest (top) bitscore, and all the matches within this threshold are reported as hits for a given query sequence. Under the Fun GCAT program, we are exploring ways to most effectively combine DIAMOND with BLASTN or other nucleotide-based classification tools like Centrifuge to optimize the timing and accuracy of results. In the current SeqScreen sensitive mode, BLASTN and BLASTX results are combined as follows:

[BLASTN results](https://github.com/shahnidhi/outlier_in_BLAST_hits) are processed through outlier detection ([https://github.com/shahnidhi/outlier\\_in\\_BLAST\\_hits](https://github.com/shahnidhi/outlier_in_BLAST_hits)) to detect which of the top hits are significantly relevant to the query sequence. The default parameters are set so that if a cut is made, all hits above the cut line are returned; otherwise, all hits are returned.

All hits within the outlier detection cutoff (BLASTN) or within 3% (default parameter cutoff=3) of the top bitscore will be saved as the top hits for a given query sequence.

Next, all hits reported by BLASTN and BLASTX are sorted by bitscore and listed for a query. Taxonomic IDs are ordered so that BLASTN are reported first, followed by BLASTX.

Order-dependent taxonomic assignments will then be based on the first taxonomic ID reported (typically BLASTN hit).

Default E-values (--evalue) and max target seqs (--max\_target\_seqs) for BLASTN and BLASTX are set to 10 and 500, respectively. Since both parameters limit the number of matches to the query sequence, modification of these parameters may be necessary for short and ubiquitous sequences. The SeqScreen DIAMOND and BLASTX protein databases were created by parsing UniProt to exclude TrEMBL and UniParc proteins. The functional annotation performance of SeqScreen improved by removing these under-annotated proteins, which typically were not reviewed or had only a 1 or 2 (out of 5) UniProt annotation score.



Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 2-8

While we are not suggesting other approaches, should the revised Guidance suggest other approaches, it would be important for the Government to make these available at no or low cost, and/or subsidize commercial offerings, or better yet, to provide API access to a screening service based on the preferred approach.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 2-9

It is the biological function of a sequence (i.e. its potential to cause harm), rather than its taxonomic origin, which is ultimately the cause for concern. Although it will not be easy to maintain a catalog of potential biological routes to harm, this is an important effort which the Guidance would ideally encourage. One effort along these lines is the Common Mechanism to Prevent Illicit Gene Synthesis, a project of the Nuclear Threat Initiative to establish “[a common, globally accessible, and regularly updated mechanism to screen nucleic acid synthesis orders and customers](#)” is an ongoing project that I expect to result in improved bioinformatics tools to address this problem. More information available here: <https://www.nti.org/analysis/articles/common-mechanism-prevent-illicit-gene-synthesis/>

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 2-10

With regards to screening of oligo pools, we encourage that the guidance also explain that the 'best match' approach is not appropriate for individual, shorter DNA sequences because of the high false positive hit rate. The guidance should recommend the use of de novo sequence assembly strategies, derived from next generation sequencing analysis approaches, as one way to estimate whether a pool of oligonucleotides could be used to assemble a gene-length fragment. Only when this approach detects a potential contiguous assembly should the sequence be subject to 'best match' sequence screening.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

## 3-1

### **Sequence Screening - Question 3: Are there other approaches (e.g., predictive bioinformatics tools) that could be utilized to identify sequences of concern for follow-up screening?**

---

While there are predictive bioinformatic tools coming online, the majority, if not all tools, lack explainability, which is required for both meaningful follow-up between synthetic dsDNA providers and their customers, as well as investigators. The issues associated with such tools stem from the fact that they rely heavily on machine learning approaches. While these approaches can provide benefit, their drawbacks are numerous. Besides explainability, one of the biggest challenges associated with these predictive tools is making sure that the training and validation datasets contain sequences with sufficient variation, so as to not train and validate these models on the same sequences. In addition, given the limited statistics for sequence types, the similarity between sequences of the same type, and poor or mis-annotation in public repositories, could result in the models making highly inaccurate predictions when challenged with a sequence for which the model has not yet been exposed.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 3-2

Not to our knowledge. There may well be tools that can help identify subsequences spread out through / embedded in larger sequences or orders (including DNA oligo pools) that could be assembled into sequences of interest; or, potentially tools that could prospectively identify hosts for the synthetic DNA that could be of concern from an emergent properties perspective (e.g. a subset of the genes required for the emergent property are present in the host genome and do not need to be supplemented in the synthetic exogenous DNA). From a chemical/small molecule threat perspective, there may be software related to retrobiosynthetic pathway design (using for example generalized chemical operator approaches) that could help assess whether a set of enzymes might possibly have an emergent property of concern (e.g. reconstitute a certain percentage of a pathway to a behavior or molecule). For the short term, much of the tooling will probably facilitate/enable/accelerate searching of known biology to identify known threats rather than to predict per se new biology or new threats.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 3-3

While a number of predictive tools are under development, we are not aware of any that are sufficiently mature to be considered a reliable standard for screening.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 3-4

Given the lack of annotated sequences in the public domain, sequence homology in amino acid space is still the most ideal metric for determining the identity of a sequence requested for synthesis. The fact that more sophisticated machine learning-based approaches simply cannot be used to address sequence screening challenges (because there is no curated public training data) is a major cause for concern among providers - and we recommend the HHS and other federal departments and agencies in the U.S. government recognize this lack of metadata around sequences that can 'endow or enhance' pathogenicity as a critical gap and undertake efforts to produce and maintain this data, akin to other such public data sets that power cutting edge science like ClinVar, OMIM and the Virus Variation Resource.

Additionally, any algorithm attempting to predict sequence of concern or other hazard must be interpretable. That is, a provider must be able not just to receive a hit/non-hit response from such a model but also to have the model provide an explanation as to why the model believes a given sequence may be subject to regulatory control. This supporting data is crucial for providers in talking with customers during follow-up screening: providers must be accountable to customers to explain why the provider is concerned about an order. Black box, 'deep learning'-style prediction approaches that do not provide interpretable answers are inappropriate for this use case.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---



## 3-5

Elimination of non-virulence associated genes encoded by bacterial and fungal select agents from the list of sequences of concern will help to reduce the number of false positives, hence lower the cost and effort of screening. While any serious pathogenic virus gene or sequence should be screened (the same is true for toxin gene sequences), most of the genes encoded by bacteria on the select agent list are not of concern. For all bacteria, only a small fraction of genes encode toxins or virulence/pathogenicity associated proteins. In the databases comprised of genome sequences of select agent viruses and bacteria, most of the genes encode proteins that pose no risks. While most bacteria are not well studied, that is not the case for select agent organisms. We propose that committees of bacterial pathogenesis experts and protein function experts for each bacterial select agent (or groups of agents) be convened to identify genes encoded by those microbes that are of potential concern. Such a process, which could be done quickly even for bacteria with large genomes, would likely eliminate >90% of the genes encoded by most if not all bacterial select agents. Removal of innocuous genes from screening lists should greatly reduce the occurrence of false positives.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

## 3-6

please see above. A value of SeqScreen is that it uses functional prediction algorithms and is thus able to detect potential threat in novel sequences.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 3-7

Yes, and our team would be happy to speak with HHS or anyone else interested in using our SeqScreen pipeline for identifying sequences of concern for follow-up screening. We also have another software product, Sequence to Functional Analysis of Threats (S2FAST), that uses SeqScreen results to make customized threat assessments. This tool could be customized to make threat determinations based on specific use cases or community guidelines.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

**3-8**

(Same as above.)

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 3-9

In addition to the Common Mechanism under development, Battelle's ThreatSeq program (more information here: <https://www.battelle.org/commercial-offerings/industry-solutions/threatseq-dna-screening-web-service>) as well other work funded by IARPA's [FunGCAT](https://www.iarpa.gov/index.php/research-programs/fun-gcat) program (more information here: <https://www.iarpa.gov/index.php/research-programs/fun-gcat>) are approaches that could be used. There are additional approaches needed for orders of smaller oligos, especially as these orders could potentially be spread amongst multiple providers. Diggans and Leproust (<https://www.frontiersin.org/articles/10.3389/fbioe.2019.00086/full>) [propose using pseudo-alignment of small k-mers](https://www.frontiersin.org/articles/10.3389/fbioe.2019.00086/full) in the near term, with longer-term effort to develop homomorphic encryption methods to allow different providers to pool orders.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 3-10

Several bioinformatic tools have been developed with the goals of predicting sequences of concern. Several of these tools, however, are based on phylogenetic conservation in amino acid identity (eg. PROVEAN, SIFT, EVMutation, PolyPhen-2). While many of these tools can, in a broad sense, predict functional damage, they are likely most useful when used together in an ensemble to predict whether a variant will damage the function of proteins. It is worth noting, however, that identifying sequences of concern might include variants that enhance function of the resultant protein (eg. higher binding affinities, higher catalytic efficiency, higher stability). These might not be effectively predicted if the mutations occur at poorly conserved residues, which these algorithms risk underestimating the impact of. For a better picture, MAVE (Multiplexed Assay of Variant Effect) databases (eg. MAVEdb) might be helpful as they are experimental information; however, I do not know if many studies have targeted pathogenic proteins.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 4-1

**Sequence Screening - Question 4: Are there other considerations that would be appropriate (e.g., batch size) in decisions about whether to conduct follow-up screening, such as oligonucleotide orders in quantities that indicate they are intended for use in assembling a pathogen genome directly?**

---

There are several technical factors that are important to weigh when considering the inclusion of oligonucleotide orders within the Guidance. From our experience developing bioinformatic tools, the minimum sequence length for which one can on average acquire species-level resolution for a single sequence's taxonomic origin is approximately fifty base pairs. As sequence length decreases from fifty base pairs to twenty base pairs, the average taxonomical resolution for a sequence can approach the domain to kingdom-level. Thus, as sequence length decreases below fifty base pairs, ambiguity on its taxonomic origin increases, making it challenging to determine the oligonucleotides intended target, which—depending on the suggested guidance—could create a false positive surplus. Coupling this surplus with the substantially higher volume of oligonucleotide orders will likely result in an inequitable situation for the industry.

As the question suggests, to alleviate the information bottleneck associated with a single short sequence, one could instead analyze batches of oligonucleotides. However, the information gain when screening batches of short sequences only works under ideal conditions; namely when the batch of sequences are from a single source for a singular use case. Given current lab practices, especially in academia, these assumptions can fail for several obvious reasons (e.g., a grad student working on several experimental projects or a lab member ordering oligonucleotides for a co-worker). Thus, as a first step in reliably solving this problem, one would need to develop a bioinformatic technology that could detect when the oligonucleotides are associated with multiple sources or use cases and group the oligo nucleotides appropriately – we suspect this problem to be solvable under certain conditions, however, further research will be required.

If such technology were to be developed that works with extremely high fidelity (recall oligonucleotide order volumes are substantially higher than gene orders, resulting in amplification of error rates), then it would likely be possible to screen oligonucleotide orders with reasonable fidelity. For example, we have tested analyzing oligonucleotides found in the methods sections of scientific papers through our next generation bioinformatic tools and found that with minimal manual review one could predict the title of the paper from a batch of oligonucleotides as small as six sequences (all of which were less than twenty

base pairs in length). While these results are promising, implementation in an operational biosecurity setting would require the automation of the manual review component to predict the use case. We believe such implementation is a solvable problem and is working to secure funding to address this challenge.

Assuming these challenges are addressed, we believe the computational resources of today would be sufficient for high fidelity screening of oligonucleotide sequences, even considering their substantially increased volume relative to synthetic gene sequences.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Orgnaization

---



## 4-2

It is not clear what batch size has to do with anything directly. Should sub-sequence embedding (either within the same sequence or across sequences) guidance be given (see above), this would address the assembly of the smaller sub-sequences into larger sequences (above the minimum sequence length cut-off). It is also not clear why a single gene hit to a pathogen genome would not trigger as much follow-up screening as hits to the entire genome. It would just be more difficult for the end user/customer to justify the order of the entire genome, rather than a single gene (which may not have any known implications in pathogenesis). Should oligonucleotides be placed into the revised guidance (see above additional forms of nucleic acids beyond dsDNA), that would cover that aspect as well.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 4-3

The Guidance should include a discussion on the use of synthetic DNA fragments for assembly into longer constructs - and that providers should monitor 'best match' species indications both a) across sequences within an order and b) across orders from the same customer. For example, a U.S.-based customer may place an order for a fragment of the genome of an FSAP-listed organism. Since this fragment cannot 'transfer pathogenicity', and as long as follow-up screening discussions indicate a reasonable intended use, it would be appropriate for the synthesis provider to make and ship the fragment. If this customer, however, repeated this process with different segments of the same viral genome, the synthesis provider will have ended up providing a full viral genome even though none of the individual orders crossed a particular regulatory threshold. The Guidance should recommend that providers track such orders within a customer over time and trigger more detailed follow-up screening if a customer looks to be attempting to acquire a full genome a piece at a time.

The Guidance should further specify that there is no way to determine solely from the number of oligonucleotide sequences ordered whether a customer intends to use these oligonucleotides to assemble longer sequences that may otherwise be subject to regulatory control. Instead, providers need to evaluate pools of oligonucleotides via de novo assembly methods to determine if assembly is a possible intended use. Methods for this assessment should take into account whether favorable thermodynamics are present for assembly and whether the overlaps in the pool are sufficiently specific to produce a final assembly product. For example, in a pool of 1,000 oligonucleotide sequences, 12 of those oligos may assemble into a controlled sequence, while the rest of the oligos are present only to obfuscate the true purpose of the pool.

DNA oligo pools have recently been used to store digital data. Because these pools are designed to last a very long time, data is stored redundantly, and the number of unique oligo sequences in such pools can be extreme: upwards of 20 million using current technology and may likely exceed 1 billion within 2-3 years. Data storage oligo pools may be ordered directly by a customer (i.e. the customer has encoded digital data into DNA sequences themselves) or the customer may provide digital data to be stored to the DNA synthesis provider who then encodes that data into DNA sequences. The former can be considered an 'untrusted' pool, since the synthesis provider had no role in its definition. In the latter model, the pool can be considered 'trusted'. For untrusted data storage pools, hiding 5-10 oligos necessary for assembly into a controlled sequence would be trivial - and screening these pools for biosecurity concerns is an area of active research at NAME. The Guidance should discuss this difference between trusted and untrusted pools and include forward-looking guidance to synthesis providers on biosecurity screening for untrusted oligo pools of all sizes. HHS should express a need for other federal departments to invest in development of efficient algorithms and tools to enable screening large oligo pools at scale for low cost.

It would also be helpful for the Guidance to provide recommendations to synthesis providers on sequence screening for variant libraries. These products generally start from an exemplar sequence and

then incorporate a very large number of sequence changes, sometimes in combination, to produce a library of variant sequences. At minimum, the Guidance should recommend that providers subject the starting exemplar sequence to sequence screening (as if it were an order for gene-length synthesis). Furthermore, if this sequence is indeed from a listed organism or toxin, follow-up screening discussions with the customer should inquire as to whether the customer has thought through any gain-of-function risk that may be inherent in the screening activities planned for the synthesized library.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 4-4

It would not be difficult for someone wishing to assemble a pathogen genome directly to attempt to disguise this among multiple orders to multiple synthesis vendors. Alternatively, a benchtop synthesis instrument might bypass detection altogether. Therefore, the new Guidance should consider order screening options that would detect attempts to split malicious synthesis orders, as well as approaches to lessen the chance that desktop synthesis instruments could be used to create known harmful effects.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 4-5

Batch sizes are worth considering in the decision tree framework for follow-up screening. Future work to the SeqScreen workflow may involve analyzing batch sizes and identifying what size of orders would be of highest priority for follow-up screening, taking into account the number of sequences required before reaching adequate coverage of sequences of concern in different pathogen genomes.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 4-6

It is important that the Guidance does not incentivize customers to distribute orders across multiple vendors-- this would be bad for biosecurity and for business.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 4-7

Yes. Diggans and Leproust (<https://www.frontiersin.org/articles/10.3389/fbioe.2019.00086/full>) note that under the [current guidance an oligo pool which could be converted into a controlled gene-length sequence would be permitted](#). They propose that oligo and oligo pool orders be first subject to a de novo assembly and the contigs be subject to a standard-length sequence screening.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

## 1-1

**Biosecurity Measures - Question 1: Is maintenance and use of broader list-based approach(es) now feasible? If so, how might this approach be realized? If not, what are major road blocks to implementing this approach? Since the release of the original Guidance, have providers or other entities developed customized database approaches, or approaches that evaluate the biological risk associated with non-Select Agent and Toxin sequences or, for international orders, sequences not associated with items on the CCL? If so, how effective have they been, and have there been any negative impacts?**

---

Yes, the maintenance and implementation of broader list-based approach(es) are now feasible (i.e., the ThreatSEQTM Web Service). However, there is a big caveat associated with this response, namely the lack of operationally trained biosecurity experts. While substantial funding has been allocated to theoretical explorations of biological risk and biosecurity practices at the policy-level, very little funding has been invested into operational

biosecurity or the training of the American workforce in a biosecurity operational setting. Thus, the only expertise that exists is present in niche groups whose cross-disciplinary workforce and required foundational intellectual property were developed over long periods of time. The investment to build both the team and the intellectual property required for effective operational biosecurity is likely too high for most biotechnology companies, as they need to ensure their profits and losses are attractive to their investors.

Further, finding and maintaining the expertise required to deliver and sustain operational biosecurity services is an additional challenge, as there are few members of the American workforce that have the experience working operationally. Thus, it behooves the Government and industry to establish a coordinated partnership and invest in programs to begin training the American workforce within an operational setting. Such a Government:private industry model should focus on, for example, the development of cross-discipline curricula, applied training models, incentive programs, and linkage of development programs to entrepreneurship, business outcomes, and social responsibility and impact.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---



## 1-2

See responses above. Combinations of best match, curated databases, and (when/where available) predictive tools should now be considered. See also below regarding sequences that do not match well to any publicly "known" sequence.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-3

In order to solve the problem of hits on housekeeping genes and to simplify screening, many have proposed a list of sequences of concern be identified, and then all orders be screened against the list, with an explicit threshold for a hit. While this system may solve the issue of hits on housekeeping genes (which we think would weaken security, not enhance it, as mentioned elsewhere), several challenges hinder effective implementation. First, challenges arise regarding who would control this system and who would have access to it. In discussions prior to the development of the current guidance, and some since then, companies did not like sending order information outside their own firewall (to protect proprietary information from their customers), so any system would have to be internal to the producer. However, entities hosting the system internally would obviously need to acquire the database, and sending the entirety of the sequence database to multiple entities would enhance the risk a malicious actor may be able to acquire it and then modify orders to avoid being flagged. Additionally, the information in any database would be of interest to hostile actors who wish to engineer a novel/modified pathogen (some of the sequences in a potential sequences of concern database may not be obvious, and seeing them present in a database may alert hostile actors to new methods or approaches to malicious use). In converse, however, tightly controlling the distribution of such a system to only known actors would pose a barrier to entry into the industry and may lead smaller firms to exit the market. Second, the use of a sequences of concern database presents another issue – there must be developed thresholds for a hit. Unlike the existing guidance, where a sequence can be compared to a "best match", if screening sequences only against putative harmful sequences, some explicit thresholds for identity or similarity must be established to use a database. Given that some microbe families that contain pathogens and near neighbors are diverse while others are extremely similar, thresholds would have to be specific for each pathogen. Diversity is also an issue while considering how best to ensure that the sequences in the database capture all Select Agents sequenced. As more pathogens are sequenced, thresholds must decrease in order to capture all variants found. Conversely, as more near neighbors are sequenced, thresholds must increase to avoid false positives. Consequently, any established threshold must be constantly and carefully re-evaluated. While this effort is feasible, it would be substantial and require ongoing updates. Lastly, and related to the above, any sequence of concern database can, by its nature, contain only sequences already identified as of concern. As a result, the database would require constant and ongoing maintenance to keep current with existing science. Certainly, the existing framework may also miss some malicious orders, but by broadly flagging any select agent sequence, a broader range of orders are more thoroughly screened (a helpful feature we believe fosters detection of a broader range of potentially malicious activities, as discussed elsewhere in our comments). If a database were to be used, we suggest it supplement and not replace the existing framework. Overall, though a sequence of concern database is a feasible component of a sequence screening system, in theory, we believe practical considerations raise serious concerns and challenges in deployment. Moreover, the database would eliminate flagging of housekeeping genes, which would hinder the ability of the screening framework to detect some types of potentially malicious activity.

Submitted on: 10/23/2020 1:19:36 PM

Agency Type: Company/Business / Agency Other:

---

## 1-4

As discussed in detail earlier in this letter on updating the scope of the Guidance, the ideal approach to sequence screening (in terms of non-ambiguity) would be for regulatory frameworks to be explicit about the sequences that are subject to control, i.e. moving from relying solely on lists of organisms (and toxins) to including lists of sequences. Even better would be to control units of biological function subject to misuse directly. This approach, however, cannot be realized without publicly available, up-to-date metadata on individual GenBank sequence records from controlled organisms that 'transfer-' (based on the criteria in FSAP) or 'endow or enhance' pathogenicity (based on the criteria in the CCL). At least one commercial screening product containing this sort of metadata now exists but the annotation, while valuable, remains proprietary to the vendor and so cannot be used by a wider community of algorithm developers to enhance screening methods nor by regulators to shape controls.

Access to sequence-level curation would also help synthesis providers significantly reduce the false positive hit rate for sequences from controlled bacteria, since the vast majority of sequences in these organisms do not 'endow or enhance' pathogenicity. Synthesis providers at present cope with this by maintaining large 'white lists' of GenBank records that have come up as 'best match' hits but, after human review, are found to have been unrelated to pathogenicity. Once added to a white list, these records can be automatically marked as 'false positives' the next time they come up during a screen and will not be subject to human review. These white lists do help to reduce false positive rates (and hence to reduce the cost of screening for providers) - but the content of white lists varies from company to company and, given that they represent significant labor on the part of each provider, companies do not generally share these white lists between companies or with the public. The Agency should consider establishing a national clearinghouse for white lists of sequences along with periodic expert (perhaps national lab) review of such white lists for accuracy. Such a resource would be extremely valuable to synthesis providers and would reduce the cost and increase the overall accuracy of screening systems.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-5

DHSS&T (under PM NAME) has developed a "Sequences of Interest" database over several years using National Lab performers. The original purpose was to provide the Synthetic Biology industry an updated gene-based database to improve order screening. However, for various reasons DHS has not made this database available to industry.

IARPA has funded the FunGCAT and FELIX programs related to screening orders and potential deliberately modified organisms, respectively. Significant software assets have been developed by these efforts, and their use could help improve how industry screens synthesis orders, if made available and leveraged properly.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 1-6

Yes, we feel the maintenance of broader databasing approaches is entirely feasible. Our team has made progress towards this under the IARPA Fun GCAT program. With our Functions of Sequences of Concern (FunSoCs) annotations and the pathogen gene ontology (PathGO) organized by the NAME, which is currently being developed under the direction of Dr. NAME, we will soon have all the conceptual tools necessary for an adequate, concise description for labeling all sequences of concern. Properly curating the sequences from all these organisms, where published data is available, is going to take some effort beyond the Fun GCAT program. During our team's performance on Fun GCAT, we have developed a screening schema using gene ontology (GO) terms and UniProt keywords to develop training sets of sequences to screen UniProt sequences with machine learning algorithms. To date, the detection of sequences of concern is not as good as it could be because of the inadequacy of the current controlled vocabularies (e.g., generic "pathogenesis" or "virulence" GO terms). With better annotated UniProt sequences using more specific FunSoCs and PathGO terms, we believe machine learning-based screening will improve and allow for better recognition for a broader array of SoCs. These would be very useful in extending the reach of all screening efforts, making automation more accurate and tractable, although completely novel sequences will still be challenging without expert human annotation.

Our software matches short sequences to databases of known functions of concern, which expands beyond the Select Agent and Toxins and items on the CCL. As the number of our expert human biocurated sequences expands with FunSoCs and PathGO terms, our machine learning algorithms also improve in their labelling accuracy because of the increased number of specific training and testing sequences available. In the threat determination outputs provided by our S2FAST software, we report the confidence and evidence underlying the threat determination for each query sequence. These general evidence categories include expert curation, taxon of low concern, point rubric, machine learning assignment, or no evidence. End users have found these evidence categories helpful to understand the basis for the automated threat assessment, which can assist in determining whether or not follow-up screening is necessary.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 1-7

Several other relevant databases exist (e.g. ThreatSeq from Batelle; SOI from LLNL,) but the Government should make them accessible and subsidize screening costs for small DNA synthesis businesses to ensure that it is economically feasible for them to follow the screening Guidance.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 2-1

**Biosecurity Measures - Question 2: Are there other security or screening approaches (e.g., risk assessments, virulence factor databases) that would be able to determine potential biosecurity risks arising from the use of nucleic acid synthesis technologies? What are the potential opportunities and limitations of these approaches?**

---

The government needs to verify that the work they are requesting of the industry promotes biosecurity and doesn't provide a "false sense of security". It should be possible to look at the screening process and give a good estimate of the potential for false negatives. This data is not available. Equally important is to have an effective screen. A screen that results in a high number of false positives is not effective. The value of the current process is not clear with the current guidance. A defined risk assessment would help industry to better assess risk and risk mitigation techniques.

Submitted on: 9/15/2020 9:16:26 AM

Agency Type: Federal Government / Agency Other:

---



## 2-2

As discussed in the question regarding predictive bioinformatic tools, explainability is a critical component of any screening approach. Risk assessments are useful for providing the Government information for prioritizing biothreats that can be used to inform the Guidance or investment. However, this type of approach is typically not useful in an operational setting, in which a practitioner needs to decide on whether to provide a specific biological sequence to a particular customer. This reality stems from the fact the provider is not driven by what would happen if they sold a wide variety of biological sequence to an ensemble of customers with varying intentions and capability. The practitioner is instead concerned with the specific biological sequence-customer pairing, and much of the strategic-level information provided by the risk assessment is not relevant to their decision.

With regard to databases, no single biothreat function database authority exists to our knowledge. Threatening functions have been identified through comparative genomic techniques and related studies leading to databases containing virulence factors, toxins, and related other sequences. However, many of these databases are incomplete, poorly maintained, and/or do not have valuable metadata for objective biosecurity assessments. Specifically, we and others have found that many of the entries in these databases simply tag sequences as "virulence factors" if attenuation of the activity leads to reduced virulence. Thus, many "virulence factors" may not be particularly threatening in the context of sequence screening. For example, the Victor's Virulence Factors Database compiles bacterial virulence factors implied from published experimentation, such as large-scale mutational screens that seek to identify attenuated virulence phenotypes. Niu et al. illustrated the controversy associated with the term "virulence factor" by determining that 69% (1,368/1,988) of virulence factors in the Virulence Factor Database (VFDB) were common among pathogens and non-pathogens.

Given such wealth of publicly available knowledge on the functions derived from genetic sequences in UniProt (and related databases), the scientific community is primed to enable function-based DNA sequence assessment. We focus our database on particularly threatening functions, which includes only a subset of virulence factor types as well as several biothreat functions not considered virulence factors. We delineate a virulence factor from a biothreat function as follows: while a virulence factor describes any factor (protein or otherwise) that aids in the virulence of organism, we define an SoC as any sequence whose verified encoded function can lead to a direct and harmful impact on a host given a biological vehicle to do so. Some traditional virulence factors are thus considered functional biothreats, such as those involved in evading the host's immune system which – when encoded in an appropriate biological context (e.g., in *E. coli*) – contribute to direct detrimental impact to the host. In contrast, a transcription factor, for example, may only indirectly impact pathogenicity, and is thus not included in our biothreat function definition. Further, we do not include factors that are typically not unique to pathogens as sequences of concern. For example, Type I and Type II secretion system proteins, which are ubiquitous throughout all gram-negative bacteria – pathogens and non-pathogens – are not

considered biothreat functions in our definition. In contrast, Types III and IV secretion system proteins, which enable transport of potentially hazardous payloads across two gram-negative bacterial membranes and a host membrane, are considered biothreat functions. More importantly, we consider several other sequence types that are not considered traditional virulence factors to be biothreat functions, such as bioregulators, animal toxins (e.g., conotoxins), protein toxins (e.g., ricin), and proteins involved in the biosynthesis of small molecule toxins and drugs.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 2-3

See above responses regarding curated databases and predictive tools. Risk assessment approaches (e.g. likelihood and severity analysis) are already implicitly if not explicitly being done, although it would be great for this process to be better specified in the guidance. Many screening softwares report out a color coded report for each screened sequence (e.g. green for sequences with no significant matches to sequences of interest, yellow for sequences matching sequences of interest but not the best match sequences, orange for sequences that are best matches to sequences on the CCL, and red for sequences that are best matches to sequences on the SAR), but this isn't a true risk assessment. Some curated databases may be doing a better job in this regard. However, should the revised guidance and associated information / databases provide a means to consistently assess risk, that would be most welcome, especially if it reduces uncertainties and labor/time associated with downstream follow up screening.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 2-4

Additional one-off data sources like the various virulence factor databases can be informative in terms of helping synthesis providers estimate the risk that a given sequence 'endows or enhances' pathogenicity. These databases, however, are generally not maintained in perpetuity and so become incomplete over time, decreasing the degree of confidence that a synthesis provider might otherwise gain by not finding a query sequence in one of these databases (i.e. the risk of false negative findings in older virulence factor data sets is too high to be useful). As previously stated, a curated database of sequences directly subject to regulatory control would be extremely valuable to providers. It is essential, however, that this database be actively maintained in perpetuity by the U.S. government up to and including same-day updates coinciding with additions or removals of organisms from the Commerce Control List. Without this degree of update, like other virulence factor databases it will age poorly and fall out of use by providers.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 2-5

Currently, the well-maintained public database of bacterial virulence factors is based in China. It would be prudent for the U.S. government to provide consistent funding to develop and maintain high-quality databases relevant for genomic screening purposes, and make them available to industry (worldwide?).

Also, better algorithms are needed that can predict functional effects of genomic editing in the context of the genome. These algorithms would need to address potential protein interactions and pathways as well as genes that are only concerns in concert.

These are two important gaps we see that are candidates for government funding to support advanced biosecurity screening.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 2-6

[PHI-base](http://www.phi-base.org/) (<http://www.phi-base.org/>) and [VFDB](http://www.mgc.ac.cn/VFs/) (<http://www.mgc.ac.cn/VFs/>) are public databases that contain virulence factors from which some sequences of concern (SoCs) could be drawn. Our team's current dataset encompasses all the verified SoCs in VFDB, as well as most of the PHI-base virulence factors. VFDB is limited to certain bacterial species pathogenic for humans, while PHI-base contains sequences from human, animal, and plant pathogens that are bacterial, viral, and eukaryotic. In neither dataset are the virulence factors comprehensively (or even mostly) annotated to describe their role in host pathogenesis. PHI-base is a regularly updated spreadsheet that contains >15,000 rows and >100 columns, many with the same or similar factors but each one reporting a different experimental situation. These datasets cannot be used for sequence screening purposes without serious and time-consuming modifications. Further, the data within them is not adequate for assessing how dangerous a particular sequence might be from the perspective of illicit bioengineering.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

## 3-1

**Biosecurity Measures - Question 3: Given that nucleic acid sequences not encompassed by SAR and the CCL may pose biosecurity risks, are there alternative approaches to the screening mechanism that could be established? If such approaches have been established, how effective have they been, and have there been any negative impacts?**

---

See responses to other questions

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 3-2

As described above, supplementing the best match approach with curated databases (beyond SAR/CCL) and predictive tools is helpful. A negative aspect of curated databases and predictive tools is that there may be a tendency to underestimate the hazard of sequences not in the curated databases or not predicted by tools. See below, for example, for discussions of sequences that do not match well to any publicly “known” sequence.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---



### 3-3

Any sophisticated approach to risk estimation must be driven by high-quality annotation of known biological functions associated with specific sequences. The synthesis providers are not well-positioned to produce this data internally as it requires expertise across multiple mechanisms of pathogenicity as well as viral and bacterial pathogenesis. The field would be best served if HHS would coordinate with others to draw on the biodefense and national lab communities as well as academic expertise across the nation in assembling such a resource.

There is also a distinct lack of annotated data around the emergent properties of collections of (otherwise unregulated) sequences that, when assembled into pathways, may be capable of causing some form of harm. For example, enabling a microbe to produce unregulated quantities of immune signaling molecules in situ in the human gut. DNA synthesis providers at present often have only a 'soda straw' view - seeing individual DNA constructs without an understanding of the larger context in which a customer will put these sequences together. A tool capable of estimating harmful outcomes from collections of sequences would be valuable to power screening at higher levels in the synthetic biology value chain, e.g. at companies specializing in organism engineering or process scale-up.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 3-4

The IARPA FunGCAT (Functional Genomic and Computational Analysis of Threats) program has resulted in excellent new functional predictive screening software such as SeqScreen. Check with IARPA regarding their experience in real world usage.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 3-5

It is worth noting that simple knowledge or a database of sequences of concern is not sufficient to create an effective bioweapon. A vast range of tacit knowledge and laboratory skills would be necessary to effectively construct a weaponizable biothreat. Although knowledge of how the sequences work together would not be necessarily present in a database housing sequences of concern, it would be critical for the design of a bioweapon. For these reasons, we do not think a list, however detailed, of sequences of concern could pose a serious threat for illicit bioengineering. However, proactive measures can be taken to protect sensitive information. When our team internally curates sequences, we include information regarding what host proteins are affected and the citation for the article in the PubMed database. Reporting on why a sequence is of concern to the user of our SeqScreen software does not require that these citations be displayed, only that the Functions of the Sequence of Concern (FunSoCs) be listed for that sequence. Similarly, when the terms are annotated with the PathGO controlled vocabulary, the citations supporting that annotation do not need to be displayed, only the PathGO term. By restricting the display of citations and the passages descriptive of the annotations, most of the utility of the database in assisting with the design of bioweapons would evaporate, while leaving the descriptive utility for end user screening intact.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

## 1-1

**Customer Screening - Question 1: What, if any, mechanisms for pre-screening customers or categories of customers for certain types of orders, if any, should be considered to make secondary screening for providers of synthetic oligonucleotides more efficient?**

---

From a security standpoint, this is the area of the guidance that has the greatest potential to really help improve security throughout the industry. For example, the primary screen could be on the customer and secondarily on the sequence. Control of access to higher risk sequences - such as large batch sequences would be more effective. Emphasis on improvement to customer screening and access would broaden security responsibilities across the supply chain, so that the providers are not only responsible.

Submitted on: 9/15/2020 9:16:26 AM

Agency Type: Federal Government / Agency Other:

---

## 1-2

There are at least several aspects to this question. The first one - general pre-screening - could be analogous to the TSA pre-check system employed at U.S. airports. For example, have the government (or some entity), perform (on a recurring basis) interviews/background checks into entities (e.g. companies, institutions) as well as individuals. While everything (i.e. sequences) would still be screened, the screening process itself (including downstream follow up) may be done in a slightly less intensive/intrusive way (e.g. no need to take off your shoes). There should still be random spot checks, just as TSA does. The second aspect is curating a repeat/returning customer that in the past has requested sequences of potential interest and has gone through a follow up process that identified that the end user/customer (or employing entity) has permits/facilities to work with certain types of sequences of concern. This could be done proactively, but at least should be done retroactively for repeat customers so that redundant follow up screening (e.g. to check on permits/facility capabilities) is not repeated. Customers should be screened for primary and secondary affiliations.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-3

In order to reduce the burden of screening, many have proposed a customer "whitelist" be developed. As proposed, this list would include customers that are known in the industry and/or that have been working without incident, such as those with a Select Agent license, Government laboratories, and large pharmaceutical companies. Proposed methods for establishing the whitelist include registration with the CDC, USDA, or even the FBI. In this scenario, whitelisted customers would not be subject to sequence screening. Developing such a whitelist would require careful consideration of the kind of information to be included in it, and the degree to which customers would have to be vetted before being whitelisted. Deployment would also require a source of funding to operate and maintain the database. When we spoke to stakeholders in 2007, and others more recently, gene synthesis companies reported that most orders are from repeat customers, so this system could greatly reduce the burden of screening by industry, though it would shift the burden to whomever is responsible for vetting whitelisted individuals. However, a whitelist is problematic to implement for foreign customers because of the need for an export permit for DNA that confers pathogenicity to a listed agent. Foreign customers must still be screened to make sure they are not ordering such DNA, or, even if they have an export permit exists that allows export of sequences of a particular type (for example, sequences from *Y. pestis*) screening would have to continue to ensure that the foreign entity isn't receiving DNA from a different, listed pathogen. If customer screening is burdensome, a sequential system could be used, as was one of the options discussed in a study we conducted in 2007. That is, the sequence could be screened, and if it isn't a best match to a pathogen, there is no need to screen the customer. The pathogens against which the sequence should be screened should probably be more diverse than the Select Agents, and include "any" pathogen. (No one would want someone to ship cDNA from measles virus to a 5th grade class). Lastly, we note that any customer whitelist may raise risks of insider threats. A benevolent industry participant that turns malevolent may bypass screening if that person is already whitelisted. Overall, we view customer whitelisting as a trade-off solution that would benefit the synthetic DNA industry, while creating costs and complexity for some other entity (likely governmental) and increasing the burden on individual researchers and industry participants seeking to be whitelisted. We also note that the whitelist would not affect the need for screening for foreign orders, and so companies would still be required to screen at least some of their customers. To reduce burden of screening, perhaps the guidance should allow sequential sequence screening (with a broader agent list) and avoid customer screening if the sequence is of no concern.

Submitted on: 10/23/2020 1:19:36 PM

Agency Type: Company/Business / Agency Other:

---

## 1-4

When a potential sequence of concern is detected, the reason for concern can typically be associated with a particular taxonomic category of threat. The cost of screening could likely be significantly reduced by allowing customers who plan to work with an organism of concern to register that plan in advance of making orders. The screening typically done for follow-up purposes could then instead be done in advance, to qualify the customer to receive sequences associated with their registered taxa. All screening hits associated with the registered taxa could then be ignored, and follow-up screening be required only for hits outside of the registered taxa.

This would not have a major impact on costs at the current rate of false positives, in which false positives dominate costs. Emerging higher-precision screening systems, such as the signature-based screening discussed above, would shift the ratio in favor of true positive hits, however. In this case, pre-screening would likely be valuable for reducing the burden of compliance with the guidance.

We do not have any specific recommendations with respect to how to implement customer pre-screening mechanisms.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 1-5

Customer screening remains a serious challenge. We recommend that the Guidance move to use the term 'restricted party screening,' a term commonly used in the logistics and global trade compliance communities, to refer to the process of checking customer names and institutions against up-to-date sets of government- and other watch lists. Such restricted party screening (RPS) tools have the unpleasant combination of high false positive rates versus exceedingly low true-positive rates (as it is truly rare for an entity on one of these lists to place an order with a DNA synthesis provider). While it is still critical to perform RPS, the Guidance should also clarify when and how often this form of screening should be performed and, importantly, how often it should be repeated as names are added and removed from various government lists. Trade compliance best practices include carrying out RPS during the initial customer onboarding as well as before any shipment. Names newly added to government watch lists should trigger re-screening of all customer and institution names against these new entries.

While RPS is critically important to ensure providers do not sell to entities already listed, it is also important for providers to know and understand the publication history, capabilities, and intentions of the non-listed customers they sell to every day. Assessing the publication history of customers after they place an order for sequence that generates a 'red flag' is a slow and laborious process - any clarifying guidance or even development of summary tools by the HHS would go a long way to harmonizing how this evaluation is carried out across synthesis providers. Even basic questions like 'has this person been employed by this institution for a significant amount of time?' can sometimes be extremely difficult to answer from public information.

Some discussion has occurred since 2010 about the relative value of 'white lists' of customers - i.e. customers with an extended business relationship with a synthesis provider or customers who have themselves been subject to some form of third-party biosecurity certification. Such customers would then, notionally, be able to order sequence without sequence screening or under some form of expedited sequence screening. We urge the government to reject this idea and for the Guidance to recommend that all customers are screened each time an order is placed or shipped. This follows trade compliance best practices in restricted party screening as lists of denied parties can change frequently.

The Guidance should further clarify that synthesis providers ask ordering customers to provide the name of the actual end user - not just e.g. the name of the principal investigator or other leadership figure in an organization. This ensures the synthesis provider has full traceability to the end user of a given synthesis order rather than just seeing the order as being sent to a large lab or other group inside a parent organization.



HHS should also propose a high-level risk assessment framework in the updated Guidance, if only to try and provide a degree of uniformity across synthesis providers in the way in which providers assess risk after follow-up screening. Such a framework could also consider providing example questions to ask customers along with a few sets of answers representing various scenarios (e.g. a legitimate researcher who is happy to talk vs. a legitimate researcher who is irritated with the synthesis company for asking these kinds of questions vs. someone actually trying to acquire material for misuse) to help elucidate when a customer should be flagged.

Provision of export-controlled material to foreign nationals on U.S. soil can be considered a 'deemed export' - but it can be challenging for a synthesis provider to determine the immigration status of an ordering party. The 2010 Guidance makes no mention of this challenge, which leaves synthesis providers with nothing to point customers to in assuring them that the questions we are asking are intended only to ensure compliance with US export regulations.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-6

Scientific users could be pre-certified to order subsets of the Select Agents and Toxins List and CCL that correspond to areas of active research. Pre-certification would be limited to only those members of a research group specified by the PI or team leader and that individual would receive email notification of all orders cleared under the pre-certification.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

## 1-7

It would help if the government maintained and authorized access to a database of companies/universities who were funded or otherwise authorized to work on specific pathogens and/or mechanisms (of virulence, toxicity, resistance, or host immune evasion, etc.)

On the flip side, it would be useful if government could assist with development and maintenance of a biotechnology relevant 'banned' list. This would include researchers that have suspected malicious intent. This project may require intel input, and would need to be updated regularly to have value.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 1-8

Screening customer orders is a critical part of the biosecurity process for commercial synthesis providers. The U.S. Government could assist commercial companies by maintaining a list of legitimate entities performing research on specific pathogens, as well as individuals or organizations banned from performing such research, and sharing those lists with authorized parties.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 1-9

The Government should offer a customer screening service for free to small businesses to ensure that it is not economically burdensome for them to follow the Guidance. Every order from every customer should still be screened. Nevertheless, we could support an approach analogous to TSA pre-check, where the extent and intensity of follow-up screening of any particular order may be dependent on this pre-screening status.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 1-10

We encourage an updated guidance document to advise against the use of so-called customer “white lists”. As you are aware, this practice involves exempting customers with an extended business relationship with a synthesis provider or customers who have themselves been subject to some form of third-party biosecurity certification from screening. The lists of denied parties can change frequently and we feel that it is best practice to screen all customers each time an order is placed or shipped. In keeping with this practice, the guidance should clarify that tool and product providers ask ordering customers to provide the name of all end users. As an example, customers should provide more than just the name of the principal investigator (especially if the laboratory that the investigator is overseeing is large) or other leadership figure in an organization. While white lists are currently impractical due largely to the open nature of much of biological research, the use of a TSA-PreCheck model may be beneficial. In this model, the level and stringency of follow up required on certain orders can be adjusted by a centralized authority and updated lists can be shared on a periodic basis. This would represent a middle ground between the full screening of every order and the carte blanche implied by white lists. We would recommend “PreChecked” individuals are screened, trained, and credentialed in a manner that is uniform and recognized by both government and industry.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

## 2-1

**Customer Screening - Question 2: Are there additional types of end-user screenings or follow-up mechanisms that should be considered to mitigate the risk that synthetic genetic materials containing sequences assessed to pose biosecurity risks are transferred to a second party who does not have a legitimate purpose to receive them?**

---

No Response

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 2-2

This will probably be very challenging to do for sequences not on the CCL or SAR (for which entities have legal/regulatory obligations to not further distribute the materials). It seems unlikely (and perhaps unjust/unethical) that one could predict the likelihood that someone or some entity would further distribute the materials. It may be possible to determine retrospectively, via provenance tracking, how sequences of interest ended up at a second party. Were this the case, perhaps that individual or entity could lose their preferred (if established) TSA-precheck status (see above) or be flagged for increased scrutiny going forward. Given this consequence, however, it seems that entities and individuals are disincentivized to track or report their transfer of these non-CCL/SAR sequences to second parties.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---



## 2-3

The global nature of the DNA synthesis market often means provider companies work with local distributors to assist with marketing and importation into other countries. The Guidance should be expanded to include explicit recommendations to ensure that both distributors and their end-customers are still subject to customer screening and follow-up screening in the case of a red flag sequence order. The Guidance should also clarify how providers should handle complex requests for drop shipment and other non-standard routing of shipments through third parties to minimize the risk of diversion.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 2-4

Guidance should promote the inclusion of customer agreements precluding transfer of sensitive materials to third party entities. It should also include mechanisms for end-users to report adverse events from use of gene synthesis or genome engineering products.

One possibility for desktop synthesizer and/or desktop genome editing instruments is that they only operate if connected to the internet (or at least connected to the ecosphere of the company providing the instrument) so that their current IP address can be confirmed to prevent unauthorized transfer of an instrument to another network. Written authorization from the instrument purchaser of record would be required to change the permitted IP address of the instrument. This would help deal with resale, theft, or other circumstances that could lead to potentially malicious unauthorized usage.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 2-5

As founders of an early-stage DNA synthesis startup, we are passionate about avoiding misuse of our DNA products, but we are also concerned about the burden of performing deep screening of our customers, and we don't have the means to investigate who our customers share their DNA with. This could also be impossible due to confidentiality agreements. This type of investigation should be the job of the Government.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 1-1

**Minimizing Burden of the Guidance - Question 1: Does implementation of the current Guidance unduly burden providers of synthetic dsDNA? If so, how could it be modified without compromising effectiveness?**

---

If the guidance does not effectively address security concerns, it is an undue burden.

Submitted on: 9/15/2020 9:16:26 AM

Agency Type: Federal Government / Agency Other:

---

## 1-2

If the current Guidance is modified properly, likely no. As stated previously, the fundamental issues associated with the current Guidance are: (1) there is no homogenization on the definition for the fundamental unit of control and (2) there is no rule-based criteria for database selection/curation and flagging for review. The impact of these issues is time unnecessarily spent on debating the Guidance and its implementation instead of on the implementation of clear and concise Guidance; hence unduly burdening the providers.

One way to efficiently and effectively implement the Guidance is through a centralized screening framework. Primary benefits would include open-access to screening tools (not privately held data); eliminated judgment-call liabilities associated with non-standard definitions, data, and methods; and the reduced cost burden of implementing, operating, and maintaining an updated DNA screening capability, which is typically a general and administrative (G&A) burden on synthetic DNA vendors. Further, by centralizing the activity, all stakeholders stand to benefit from transparency and standardization of methods and reference data. As enrollment in the capability grows with the industry, and only with participant agreements in place, a potential cross-vendor visibility (i.e., federated screening) may offer an additional layer of biorisk management. A well-negotiated centralized model with properly funded O&M activities would afford both a scalable and sustainable solution for the rapidly growing, global synthetic DNA industry.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 1-3

No, it does not unduly burden providers. It does come at a cost, and this can be a competitive disadvantage on a playing field where there are no regulations (only voluntary self-governance) and competitors that do not comply. The solution to this conundrum is probably not modifying the Guidance as much as to (for example) institute government procurement contract requirements, etc., to incentivize compliance. Federally-funded providers could also have operational requirements for compliance that are specified and are a requirement of receiving federal funds.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-4

Biological threats have extremely high potential impact, so the burden is not undue, given the \$15/order cost reported in the JCVI 2015 report "DNA Synthesis and Biosecurity: Lessons Learned and Options for the Future."

The emergence of higher-precision screening systems, such as the signature-based screening discussed above, has the potential to greatly reduce the effective costs of implementation. To enable this, however, the guidance should be modified to remove the specification of BLAST as a required standard, since BLAST tends to produce false positives at a higher rate than other emerging technologies.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 1-5

Screening under the 2010 Guidance is an important component of ensuring responsible use of synthetic DNA. A near-complete lack of data, technology, and tool investment on the part of governments globally means that, in 2020, many smaller providers are still using the same exact tools they were using in 2010. As such, the costs associated with screening have not been reduced at all while the price per base of DNA has gone down by at least an order of magnitude. In fact, with the increase in the size of databases like NCBI's nr, the compute requirements in 2020 are larger and more expensive than they were in 2010, all else held constant. This leads biosecurity screening to represent an ever-larger percentage of the total cost of production for synthetic DNA, creating a strong economic disincentive for smaller providers to screen at all. This is not to say the Guidance should be modified substantially - to do so would weaken the value of sequence screening. Instead, it is an indicator of the potential for impact on global biosecurity (and bioeconomy growth more broadly) from investment in these sorts of new data sources, algorithms and technologies.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---



## 1-6

A 2015 report from the J. Craig Venter Institute, DNA Synthesis and Biosecurity: Lessons Learned and Options for the Future <https://www.jcvi.org/research/dna-synthesis-and-biosecurity-lessons-learned-and-options-future>, estimated the time spent and cost of bioinformatic screening synthetic dsDNA based on data supplied by IGSC member companies. Screening related costs were approximately \$15/order, about 1.5% to 3.0% of a typical order. About 13% of the cost was due to professional-level bioinformatics-staff time to determine whether the order needed further review. About 60% of the cost was for customer follow-up for orders unlikely to cause harm, i.e., false positives. About 13% of the cost was for customer follow-up for correctly flagged sequences, i.e., true positives that were being ordered by legitimate users. The sequence screening system described above is estimated to lower the rate of false positives by a factor of three. Initial screening costs would be lower as well. If combined with pre-screening of legitimate users as described above, total screening costs could be lowered to well below half of today's BLAST-based approaches.

The cost of sequence screening might be further reduced by allowing users to self-screen sequences before submitting, in order to avoid unintentional true positives. There may be a significant computational cost for doing this with BLAST-style screening, but signature-based screening (as described above) reduces this greatly. Self-screening would need to be managed carefully to prevent reverse engineering, but this risk could readily be managed with screening servers that make use of user authentication and rate limiting, just as is used to prevent password guessing in financial transactions.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---

**1-7**

It seems that a standardized dataset of sequences known to be concerning would be a great benefit to synthetic dsDNA providers. Presently, they all have individual approaches and databases to screening sequences. A standardized dataset would, in our understanding, lower both costs and certain worries about legal liability for the providers.

Submitted on: 10/25/2020 5:04:21 AM

Agency Type: Company/Business / Agency Other:

---

## 1-8

Yes, it definitely places a burden on providers of synthetic DNA, but we strongly support enhanced screening of DNA orders and customers. In order for this crucial screening to be successful, it is imperative that the Government support small businesses in the screening process of both customers and their orders. For the orders, in particular, we recommend that the Government provide an API for screening sequences that returns results in a timely manner (less than 2 hours). Providing such a service would create a level playing field for small businesses and ensure that large companies don't have an unfair advantage from greater resources available to implement screening software and processes. The screening assistance/API would also prevent a situation where companies that do not follow screening Guidance would gain a competitive advantage compared to those that responsibly follow the Guidance.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 1-9

Yes; current algorithms have a high false-positive rate and the results require expert interpretation. Providers are also burdened by developing their own in-house bioinformatic screening systems, an issue that would be addressed if NTI's Common Screening Mechanism or a similar proposal is adopted.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

# 1-10

Implementing the Guidance is expensive, and it continues to increase (with the increased provision of dsDNA, with the size of public sequence databases, with the cost of curated database vendors, with the cost of operating expenses for downstream follow-up screening).

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-11

There has been some drift in the factors that contribute to the cost of screening over the past ten years. The nr database has grown in size which increases the compute requirements (and therefore cost) for blastx-style 'best match' screening. At the same time, cloud computing resources have grown cheaper over time and newer alignment algorithms (producing the same output as blastx-style alignment but using more efficient k-mer-based search strategies) have reduced overall runtime for sequence database searches.

The rate of change in these factors, however, has flattened - they are unlikely to give us any additional advantage in keeping the costs of screening low over the next decade. Instead, we need to tackle head on the most important and most expensive part of screening: the need for human review of any red flag sequence. Providers need new annotated data resources, tools and approaches to keep the price per base of DNA declining without biosecurity becoming a leading component of that per-base cost. Specific recommendations to provide companies with these resources are made throughout our comments.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-12

I don't have personal experience with this, but Diggans and Leproust (<https://www.frontiersin.org/articles/10.3389/fbioe.2019.00086/full>) wrote that "[As scale drives down cost per base pair, the relatively fixed cost of screening plays a more direct role in overall price.](#) These costs are driven by both customer and sequence screening—commercially-available customer screening solutions still require a great deal of manual review of false positive findings. These false positives create a floor on the possible reduction in labor cost of new customer onboarding. Current sequence screening algorithms are computationally expensive and, given the high false positive rate, the results of sequence screening can be complicated to interpret. These generally require a PhD in bioinformatics both for implementation as well as day to day interpretation of hits. This makes scaling interpretation, in the absence of high-quality sequence annotation, a very expensive proposition.”.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

# 1-13

We do not believe there are any challenges associated with retaining records of customer orders, “hits”, and/or follow-up screening for at least eight years because cold storage for large volumes of data is more affordable than ever before and its cost is only projected to decrease.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---



## 1-14

No substantial/unexpected challenges. We have infrastructure now in place to do this. However, it may well present a challenge for other organizations that do not have their own extant infrastructure to do this. Commercial vendors or government supported / subsidized infrastructure would probably be welcome in this space.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-15

This recommendation does not present a challenge other than the cost of storage, considering that most modern screening platforms store these results 'in the cloud' and end up paying the accumulated storage costs for screening results over the 8-year period. With high-latency, archive-tier storage, costs for this retention are manageable compared to the cost of screening itself. It would, however, be valuable for the Guidance to provide more detail on how 'live' this data needs to be - e.g. there are archive storage strategies that increase the amount of time required to access this data (to 12 or 24 hours) while reducing monthly cost. Given the purpose of this data, it is probably entirely acceptable to store these data in 'deep archive' storage for lower cost, but having HHS affirm this would be valuable to existing synthesis companies as well as helping newer synthesis companies understand where they can reduce cost while maintaining adherence to the guidance.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-16

This appears to be the minimum that a responsible organization involved in synthetic biology (synthesis or editing) should perform. Longer timeframes for record retention present challenges related to database storage and maintenance of customer confidentiality. For a new start-up company, 8 years is a long time. and mitigation processes would need to be put into place in the event a company ceases to exist.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

**1-17**

More specifics would be helpful in identifying what information needs to be preserved. Need to specify what data we actually need to retain.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 2-1

### **Minimizing Burden of the Guidance - Question 2: Have customers experienced delays in receiving orders of synthetic dsDNA due to screening?**

Yes, certainly. Just performing the screening (even without any hits to sequences of interest and no follow-up screening required) requires some time and will delay order fulfilment (even for hours to days). Should there be any hits to sequences of interest, follow-up screening can further delay the process (days to weeks) or even result in the cancellation of the order. Such delays could be mitigated through dedicated resources being made available such as standardized and updated portals or dedicated funds for screening and ensuring compliance.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 2-2

The sequence screening process itself can be completed well before a sequence moves to manufacture within a synthesis company. Delays occur, however, once a red flag is found and follow-up screening must be performed. This results in email and phone traffic back and forth between the customer and the synthesis company support staff. Given current industry turn-around-time, as long as no export license is required, follow-up screening can generally be completed before the provider is ready to ship out ordered product.

If an export classification or license is required, however, this process causes significant delays to delivery of sequences. Turnaround time in these cases can be five- to ten times normal and uniformly results in upset customers and delayed scientific experiments. There are two primary challenges around export licensing of synthetic DNA products: 1) detecting the need for a license at the time of sale and 2) applying for a license. Given HHS's expertise, NAME strongly urges the Guidance to be updated to clarify terms, like the term 'gene', that could then be adopted by the Department of Commerce for determining when export licenses are and are not required.

Given the lack of sequence metadata discussed previously, there is no possible way to cleanly automate the detection of sequences and confidently warn a customer that a given sequence they are considering ordering may be unique to a CCL-listed entity and 'endow or enhance' pathogenicity. This process must instead occur after a customer has placed an order and it is extremely common for customers to cancel orders once we alert them to the fact that their order may require a license. They often then re-order the same sequence from providers in their home country, who may or may not employ biosecurity screening. While we continue to believe that HHS must update screening standards and that U.S. companies must adhere to these standards, it is important to note that this lack of automatability in the determination of export-controlled status results in lost business for U.S.-based synthesis companies.

As discussed in detail in the first section of our response, defining when a sequence is of sufficient length or of sufficient homology (to a known, publicly available gene sequence annotated as coming from a listed organism or toxin) to be considered a 'gene' and so subject to a license requirement would add greatly to biosecurity efforts and promote competitiveness by ensuring all companies adhere to the same licensing requirements. NAME believes that HHS has a critical role to place in that process. Without clear guidance as to what constitutes a 'gene' for the purposes of control under ECCN 1C353, the burden falls on each individual DNA synthesis company to identify and submit classification requests to the Commerce Department's Bureau of Industry and Security for what the provider believes to be borderline sequences. As stated earlier, this current practice results in: (1) increased variability in export compliance by U.S. based DNA synthesis companies (in that some companies may seek a classification while others may not), (2) extended turn-around time for sequence delivery, and (3) increased internal labor costs per base pair for U.S. based DNA synthesis companies. These in turn reduce the competitiveness of American synthesis providers globally.

Even if a customer agrees to wait for a license, the application process via SNAP-R is very manual, requiring a human to click through a web interface to file a license application. NAME operates at scale using highly automated computer systems - and these systems are more than

capable of automatically submitting license requests if the receiving U.S. government systems were capable of this kind of API-based automation.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 3-1

**Minimizing Burden of the Guidance - Question 3: Have there been any undue burdens, financial, logistical, or otherwise since implementing the Guidance? If so, has it increased, especially as other costs associated with dsDNA synthesis have decreased?**

Implementing the Guidance is expensive, and it continues to increase (with the increased provision of dsDNA, with the size of public sequence databases, with the cost of curated database vendors, with the cost of operating expenses for downstream follow-up screening).

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---



## 3-2

There has been some drift in the factors that contribute to the cost of screening over the past ten years. The nr database has grown in size which increases the compute requirements (and therefore cost) for blastx-style 'best match' screening. At the same time, cloud computing resources have grown cheaper over time and newer alignment algorithms (producing the same output as blastx-style alignment but using more efficient k-mer-based search strategies) have reduced overall runtime for sequence database searches.

The rate of change in these factors, however, has flattened - they are unlikely to give us any additional advantage in keeping the costs of screening low over the next decade. Instead, we need to tackle head on the most important and most expensive part of screening: the need for human review of any red flag sequence. Providers need new annotated data resources, tools and approaches to keep the price per base of DNA declining without biosecurity becoming a leading component of that per-base cost. Specific recommendations to provide companies with these resources are made throughout our comments.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 3-3

I don't have personal experience with this, but Diggans and Leproust (<https://www.frontiersin.org/articles/10.3389/fbioe.2019.00086/full>) wrote that "[As scale drives down cost per base pair, the relatively fixed cost of screening plays a more direct role in overall price.](#) These costs are driven by both customer and sequence screening—commercially-available customer screening solutions still require a great deal of manual review of false positive findings. These false positives create a floor on the possible reduction in labor cost of new customer onboarding. Current sequence screening algorithms are computationally expensive and, given the high false positive rate, the results of sequence screening can be complicated to interpret. These generally require a PhD in bioinformatics both for implementation as well as day to day interpretation of hits. This makes scaling interpretation, in the absence of high-quality sequence annotation, a very expensive proposition.”.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

## 4-1

**Minimizing Burden of the Guidance - Question 4: What challenges, if any, do the recommendation to retain records of customer orders, “hits,” and/or follow-up screening for at least eight years present for your organization?**

---

We do not believe there are any challenges associated with retaining records of customer orders, “hits”, and/or follow-up screening for at least eight years because cold storage for large volumes of data is more affordable than ever before and its cost is only projected to decrease.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 4-2

No substantial/unexpected challenges. We have infrastructure now in place to do this. However, it may well present a challenge for other organizations that do not have their own extant infrastructure to do this. Commercial vendors or government supported / subsidized infrastructure would probably be welcome in this space.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 4-3

This recommendation does not present a challenge other than the cost of storage, considering that most modern screening platforms store these results 'in the cloud' and end up paying the accumulated storage costs for screening results over the 8-year period. With high-latency, archive-tier storage, costs for this retention are manageable compared to the cost of screening itself. It would, however, be valuable for the Guidance to provide more detail on how 'live' this data needs to be - e.g. there are archive storage strategies that increase the amount of time required to access this data (to 12 or 24 hours) while reducing monthly cost. Given the purpose of this data, it is probably entirely acceptable to store these data in 'deep archive' storage for lower cost, but having HHS affirm this would be valuable to existing synthesis companies as well as helping newer synthesis companies understand where they can reduce cost while maintaining adherence to the guidance.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 4-4

This appears to be the minimum that a responsible organization involved in synthetic biology (synthesis or editing) should perform. Longer timeframes for record retention present challenges related to database storage and maintenance of customer confidentiality. For a new start-up company, 8 years is a long time, and mitigation processes would need to be put into place in the event a company ceases to exist.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 4-5

More specifics would be helpful in identifying what information needs to be preserved. Need to specify what data we actually need to retain.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 5-1

### **Minimizing Burden of the Guidance - Question 5: How might potential changes to the Guidance to expand the scope or methodologies affect the burden for providers of dsDNA and customers (including delays to scientific progress caused by extended review)?**

In order for the revised Guidance to have a positive impact on providers and customers, the Government needs to ensure the revisions provide: (1) a clear definition of the fundamental unit of control, (2) quantitatively define the flagging criteria that triggers order review and customer follow-up, and (3) a mechanism to offer providers access to a standardized screening system that protects providers and customer data.

Our proposed standardization along with routine, science-informed maintenance is expected to provide vendors with a comprehensive functional (and threat) assessment of synthetic DNA sequences with agreed-upon methods and systems for productive and efficient customer engagement and decision support. The proposed partnership model will establish a collaborative and accountable biosecurity ecosystem to deliver a practical, accessible, and cost-effective screening solution that meets the requirements of the evolving guidance, obviating the need to devise a more restrictive regulatory framework.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---



## 5-2

Expanding the scope and/or methodologies will certainly increase the cost and timelines for order fulfillment. The question is one of cost/benefit: are these increases in cost and delays in fulfillment justified given the prospective benefits to biosafety/biosecurity.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 5-3

The goal of the Guidance is to support the adoption of a baseline set of best practices across U.S.-based providers of synthetic DNA. Those best practices must continue to evolve along with the growth of technology and capability both within providers of synthetic DNA as well as among our customers. This maturation of screening best practices may bring with it increased cost in terms of both screening implementation as well as efforts to assess the performance of constructed screening systems. However, these changes are sorely needed, and we have demonstrated that the enhanced standards can be part of a successful and thriving business.

We note that the recommendations we make here would not appreciably increase the time required for sequence screening or follow-up so there would be no risk of delay in scientific progress. HHS could play a powerful role in reducing the cost of screening borne by providers by assembling and maintaining high-quality databases of units of biological function that can 'endow or enhance' pathogenicity from listed organisms. As discussed in response to another question, such a database would allow providers a much greater degree of automation in biosecurity sequence screening and drastically lower the frequency of false positive findings, both of which would significantly lower sequence screening costs.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 5-4

Providers will maintain their own database of “trusted customers” so subsequent orders by the same customer for the same organism should not require delays. Customers should rapidly become educated to the fact that work on potentially dangerous sequences will require them to provide explanation and documentation for the order to be filled. The Guidance should delineate what this documentation should include.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 5-5

We are the founders of an early-stage startup offering a rapid DNA synthesis service. Our value proposition is that we can synthesize and deliver long dsDNA very quickly. If screening is slow (half a day or more) it will delay our manufacturing start time and will eat into our value proposition. Slow screening will also significantly slow down the research of the end users. However, we strongly support thorough screening of orders, so the Government should provide a screening API that returns answers in 2 hours or less.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 6-1

### **Minimizing Burden of the Guidance - Question 6: Is your organization concerned about legal liability challenges between customers and providers?**

---

Yes, this is a concern for any legitimate business, especially when the goal of the Guidance is to minimize the biological risk associated with the industry (i.e., increase the time between biological events), as completely eliminating all biological risk is an intractable goal.

Our proposed U.S. Government:private partnership model will foremost reduce liability risk through routine stakeholder engagement to harmonize on definitions, methods, and reference data. To further ensure the quality of a centralized system, the U.S. Government could conduct routine technology audits and implement findings for continuous capability improvement.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Orgnaization

---

## 6-2

Yes. As well, we are concerned about doing the right thing, whether dictated by legislation/regulation or just best and sensible practices.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 6-3

Synthesis providers have often tried to address this question of liability (especially in the case of sequences known to be toxic but not regulated for possession) by requiring customers to sign a 'biosafety waiver' or other document when placing an order for a sequence with known harmful risk. The efficacy of these documents in terms of mitigating liability (if indeed any exists) is unclear and untested - and it would be valuable for the Guidance to discuss this challenge explicitly and to make recommendations in terms of best practices.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 6-4

Some customers may desire to provide the absolute minimum amount of information, to preserve their Intellectual Property related to how they modify organisms to produce useful products. The new Guidance should clearly spell out what information must be forthcoming and how the providers should safeguard that customer information while maintaining customer trust and satisfaction.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---



## 6-5

As long as we (DNA synthesis service) follow the guidance, liability for misuse should fall onto the customer.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

## 1-1

**Technologies Subject to the Guidance - Question 1: Do other oligonucleotide types and other synthetic biological technologies, currently not covered by the Guidance, pose similar biosecurity risks as synthetic dsDNA (e.g., Ribonucleic Acid [RNA], single-stranded DNA, or other oligonucleotides)?**

---

Yes, but many of these issues are easily mitigable with cloud hosted screening or fail-safes in the tools themselves.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Orgnaization

---

## 1-2

Yes. The likelihood/severity of these risks may be different than dsDNA, however, and a biosecurity risk assessment should be performed to determine severity/likelihood before including into any expanded Guidance.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 1-3

We have conducted studies on behalf of the US Government examining the how misuse of other parts of the emerging biotechnology industry could create a biosecurity risk. At the highest level, the misuse of gene synthesis companies offers an adversary the opportunity to acquire a potentially harmful product relying on the skills of others. In other words, misuse of industry enables a low skilled actor to misuse highly skilled individuals in pursuit of malicious aims. In that broader risk assessment context, any biotech industry that offers a similar capacity could also be ripe for misuse. Generally, companies offering these kinds of services can be thought of as outsourcing providers. Our research and analysis concluded that the availability, breadth, and depth of outsourcing services with biotech (specifically, synthetic biology) is growing over time and that real risks of misuse are present in some of these companies. In other words, these industries represent another category of products that malicious actors can directly misuse to achieve harmful aims.

After considering how best to reduce misuse risks, we concluded that the same two principles that shape the existing HHS guidance still hold for this larger set of companies, and to those two principles we suggest adding another:

Companies have a responsibility to know their customers

Companies have a responsibility to know what they are making/selling

Companies have a responsibility to know their legal and ethical responsibilities.

The first two principles come from the existing guidance, and apply equally to gene synthesis companies, protein production companies, strain engineering companies, and other biological "outsourcing" firms. If those companies know what they are making could be harmful, and also know who their customers are and whether they have a legitimate purpose to order anything that may be harmful, the risk of misuse can be significant mitigated, no matter the product.

The third principle stems from our engagement with stakeholder companies, and a discovery that not all of them were aware of their existing statutory requirements, available guidance, or industry best practices. For example, at least one company facilitated the transfer of a Select Agent without having heard of the Select Agent program. Another stated they were aware customers could submit orders for harmful products, but purposefully chose not to screen them. In both examples, we suggest no laws were broken, but each example highlights gaps in biosecurity present today.

Overall, we suggest that many industries beyond the synthetic DNA industry should implement a screening framework, and we suggest the US Government consider who those industries are. We also suggest that any new guidance be paired with new educational documents and/or training to industry, to help them be better aware of their existing requirements and responsibilities. Beyond new emerging industries and technologies, many life-science products have long been recognized for their dual-use potential. Long before the de novo synthesis of polio, the Australia Group was established to

prevent the international trade in items usable for the manufacture and use of weapons of mass destruction. The Australia Group list (which is imbued with the force of law in the US by its embodiment in the Commerce Control List--CCL) contains many life-science products and equipment (fermenters, spray driers, foggers etc) that could be used for the isolation, magnification and dissemination of pathogens. Because these items are found on the CCL, they cannot be exported to non-Australia Group countries without an export permit. However, there is no mechanism, guidance, or even a suggestion from the government that the trade of such items should be scrutinized between domestic suppliers and customers. Clearly, a mechanism to apply the know-your-customer rubric should be applied to the domestic trade in dual use goods to prevent the sale of a complete bioproduction facility to a malicious actor based in the US. Many of these items do have many legitimate uses, so a customer/order screening system would be appropriate. However, the Australia Group list and CCL are focused on state-actors and large-scale programs for weapons of mass destruction. If dual-use biological items are to be subject to guidance in the US, additional items should be considered. For example, smaller fermenters and materials suited for the production of kilograms (not thousands of kilograms) of biological agent (like micronizers in addition to hammer/bead mills) should be considered. Moreover, some products are ONLY used for the production of pathogens, so knowing the customers for these goods is an obvious goal with little downside (for example, heart-brain infusion broth or sheep's blood agar). Nearly anyone working with those products is working with pathogens, and measures should be taken to make sure domestic purchasers of these goods handle them responsibly. In proposing these products be added to the screening guidance, we recognize the risk of false positives, whereby a legitimate user has an order held during screening. Of note, the clinical laboratory sector commonly uses these products, and recent experience during the pandemic has illustrated that delays in the clinical lab supply chain can have a pointedly negative influence on preparedness and response. One way to reduce the likelihood of false positives in this sector would be for labs to include their CLIA registration numbers in orders, which could be easily verified against an existing database maintained by CMS and CDC.

Lastly, we share one specific point about synthetic RNA. Synthetic RNA also poses biosecurity risks because rescue platforms for some Select Agent RNA viruses involve introduction of viral RNA into cells. We note that the select agent regulations cover "nucleic acids that can produce infectious forms of any of the select agent viruses" (i.e., full genomes or genome segments, for segmented viruses), but do not cover fragments of genomes/genome segments.

Submitted on: 10/23/2020 1:19:36 PM

Agency Type: Company/Business / Agency Other:

---

## 1-4

As noted above, dsDNA can now be readily produced from ssDNA, so ssDNA should be subject to the guidance as well. RNA is not commonly used to produce dsDNA, but could be used for doing so and thus should be subject to the guidance to avoid a workaround. XNA should be considered: it does not appear mature or widespread enough to be a readily accessible source of threats at present, but that also means that the cost of compliance for XNA may be low since the market is small.

We are also recommending reduction of length from 200 bp to somewhere in the 50-75 bp range, on the following basis:

Some synthesis providers already routinely screen 75 bp sequences without undue burden.

Performers in the IARPA FunGCAT program have demonstrated that 50bp sequences can readily be distinguished with less than 1% false positives and without any increase in false negatives.

Very short oligonucleotides, such as those typically used for primers, do not generally pose a threat due to the difficulty of assembling these into larger sequences and still would not be affected by this guidance.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 1-5

As previously stated, the guidance should make no differentiation between single-stranded DNA or double-stranded DNA. It is trivial to convert between the two. This same argument extends to RNA which can be converted back to DNA with off-the-shelf reagent kits. The Guidance should make no distinction between synthetic single- or double-stranded DNA or RNA - all should be subject to the same level of Guidance-defined sequence screening and follow-up.

Also as previously stated, the Guidance should formally incorporate oligonucleotide pool screening as a necessary component of a broadly effective biosecurity screening program. These risk estimation methods for oligonucleotide pools, in particular, must be considered in the context of the use of oligonucleotides for data storage in DNA. Storage of 1 Tb of data, e.g., using a reasonable estimate of data density would result in a pool of 250 million unique oligonucleotide sequences. Ordering a pool like this could be used as a complex background in which to include oligos that could assemble into controlled sequences - raising the importance of algorithms that can efficiently generate hypotheses about thermodynamically likely assembly products even in truly enormous pools of this size.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 1-6

As discussed above, we believe that the Guidance needs to be expanded to include providers of synthetic oligonucleotides. Of course, a key consideration for screening oligos is whether costs can be held to an acceptable level. First, not all oligos would need to be screened. Oligos used for gene synthesis are typically between 40 to 60 nucleotides in length. The above-mentioned JCVI report estimated that setting a lower bound of 40 nucleotides for screening would exempt 90% or more of the oligo market (most of which is small oligos used for PCR). Using non-BLAST based approaches such as that developed by a performer in the IARPA FunGCAT program (described above) for the remaining 5 to 10% of the oligo market could lower screening costs substantially. At minimum, oligo providers could be encouraged to screen their customers to make sure that the oligos are going to a legitimate institution or known entity, even without sequence screening. U.S. oligo manufacturers already have an obligation to screen against U.S. government export control watch lists if they are shipping to other countries. (The JCVI report cited above includes additional discussion about ways to reduce the cost of oligo screening.)

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---



## 1-7

(also see responses to the scope of guidance questions above) Guidance should cover identification and evaluation of guide RNA and repair template sequences in customer orders, noting that the sequences may be in the context of a plasmid construct. Short guide sequences determine target sequences that will be edited in double strand break technologies, and repair templates provide sequence that will be inserted into edited cells.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 1-8

Absolutely-- the Guidance should cover synthesis of any information-carrying polymer that can be converted into dsDNA with good fidelity. This includes ssDNA, RNA, XNA, etc. This is essential given the improvements expected in ssDNA and RNA synthesis over the coming years.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 1-9

Yes. Since molecular biology techniques allow various synthetic nucleic acids to be transformed into one another, the screening guidance should be broadened to apply to other nucleic acids such as synthetic RNA and ssDNA. The de novo synthesis of the horsepox virus in 2018 also shows clearly that oligo pools, perhaps of less than 200bp in length, may be assembled into sequences of concern.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 1-10

We greatly appreciate that you invite comment on the current focus of the guidance on only synthetic double-stranded DNA. Single- and double-stranded DNA conversions are commonly carried out and additionally, RNA can also be interconverted to DNA using off-the-shelf reagent kits. We also recommend that HHS extend guidance to include screening of oligonucleotide pools. Providers of synthetic genetic material regularly use oligos smaller than 200 nucleotides to assemble and manufacture gene-length DNA sequences. Given the ease of performing these tasks, we strongly urge HHS to broaden the guidance to include all types of synthetically generated DNA and RNA, along with pools of shorter oligonucleotide sequences. With regards to screening of oligo pools, we encourage that the guidance also explain that the 'best match' approach is not appropriate for individual, shorter DNA sequences because of the high false positive hit rate. The guidance should recommend the use of de novo sequence assembly strategies, derived from next generation sequencing analysis approaches, as one way to estimate whether a pool of oligonucleotides could be used to assemble a gene-length fragment. Only when this approach detects a potential contiguous assembly should the sequence be subject to 'best match' sequence screening.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

---

# 1-11

RNA, ssDNA, etc might pose risks since they can be duplexed and reverse-transcribed as needed.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 2-1

**Technologies Subject to the Guidance - Question 2: Are there other appropriate security measures that should be established to address the potential threats arising from the use of nucleic acid synthesis, given new and emerging technologies in the life sciences?**

---

See other responses.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 2-2

See above regarding the Guidance also applying, for example, to software infrastructure in addition to the point of physical fabrication.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 2-3

The 'know your customer' concept spelled out in the 2010 Guidance is an extremely valuable normative practice that should be recommended in the Guidance to apply to the entire synthetic biology value chain, not just to the providers of synthetic DNA. At any step in this value chain, a company should be screening its customers against lists of denied parties and determining whether a customer has the kind of prior work that would properly frame the material they have requested (organism, protein, DNA, RNA, etc.). Any company with a concern about mismatched expectations or misrepresentations should feel comfortable taking these concerns to their local FBI WMD coordinator. If the Guidance is narrowly focused only on providers of synthetic DNA, these additional companies may feel explicitly that they have a diminished responsibility -- that the DNA synthesis providers are 'taking care of' these risks.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---



## 2-4

The new Guidance should recognize that whole genome editing technologies may involve massively parallel edits (each in different cells or arbitrary combinations in the same cells.) These complex changes need to be evaluated in context, unlike simple oligo orders that can be examined individually. Screening of proposed complex edits can be accomplished during the design phase, meaning that potentially risky edits can be discussed with the customer or excluded from the order prior to reagent creation.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 2-5

Yes, we strongly believe that benchtop DNA synthesizers pose a serious biosecurity threat and that access to them should be tightly controlled. No amount of DRM or other security engineering will be sufficient to completely prevent hacking, “jailbreaking”, and other misuse, intended by the user or performed maliciously by a third party, with a device that simply dispenses a handful of liquid reagents into a plate repeatedly. The economic and societal damage that can be unleashed by the spread of a novel pathogen is particularly evident right now due to the COVID-19 pandemic. We believe that easy access to DNA printers would significantly increase the likelihood of similar or worse events in the future based on engineered pathogens. One could argue that oligonucleotide synthesizers have been commercially available for decades now without causing serious biosecurity incident, so no new regulations are needed. However, this is not a valid argument because 1) the capabilities of new benchtop DNA synthesizers will greatly exceed those of previous synthesizers in terms of speed, fidelity, scale, and ease of use, and 2) our ability to engineer biological systems has significantly increased in recent years due to the advent of new DNA manipulation technologies such as Gibson Assembly and CRISPR-Cas9. Therefore, access to new benchtop synthesizers must be restricted to reduce the risk of misuse.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 2-6

In order to minimize the likelihood that synthetic biology products and tools are used for nefarious purposes, we highly recommended that the guidance be expanded to include recommended customer screening practices for all providers (e.g. organism engineering, genetic circuit design, protein engineering firms, etc.) in the synthetic biology supply chain. This would bolster protections and convey that all product and tools providers have an equal responsibility in ensuring that these items do not end up in the wrong hands. All companies should be screening its customers against lists of denied parties and determining whether a customer has the kind of prior work that would properly frame the material they have requested. Furthermore, list of denied parties should be consolidated and shared amongst providers in an origin agnostic manner. Additionally, any company concerned that a customer may be considering using a product inappropriately should be encouraged to discuss these concerns with their local FBI WMD coordinator. Sequence screening is a valuable tool for limiting the misuse of engineering biology and associated technologies. It should be incorporated into a holistic security strategy. Another part of that strategy should be the collection HUMINT from customer screening, newly identified sequences of concern, as well as security inputs from other sectors (e.g., names of those attempting to inappropriately acquire other potentially destructive materials); harmonizes these materials; and communicates concerns back to stakeholders in the biotechnology (and other) sectors.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

## 3-1

**Technologies Subject to the Guidance - Question 3: Are there new biosecurity risks posed by the introduction of new generations of benchtop DNA synthesizers capable of synthesizing and assembling dsDNA, RNA, single-stranded DNA, or oligonucleotides in-house that should be addressed by the Guidance?**

---

Yes, but many of these issues are easily mitigable with cloud hosted screening or fail-safes in the tools themselves.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 3-2

As long as it is clear that benchtop DNA synthesizers and related technologies are also following the guidance (e.g. they phone home), for many types of risks, the point of fabrication should not impact risk (severity/likelihood) as long as the quality of the product delivered is comparable to that prepared in-house. (For example, some reagents are too unstable to be prepared externally and then shipped, so certain risks may only be applicable to in-house benchtop synthesizers). While there could be malicious attacks on the hardware/instrumentation in the synthesis and assembly process, whether at a centralized facility, or across many distributed sites (via for example benchtop instruments), the distributed nature of the benchtop synthesizers may present different security challenges than the instruments in centralized facilities. For example, there may be different levels of cyber and physical security protocols in place across many distributed facilities operated and controlled by distinct entities. Thus identifying security risks in a distributed context may be more complex and more difficult to control (for example by the manufacturer).

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 3-3

As noted above, we recommend that the guidance be extended to benchtop synthesizers. We also recommend that the guidance be extended to other benchtop systems, such as benchtop editors.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 3-4

The guidance absolutely must address the risk posed by benchtop DNA synthesis devices - a benchtop device capable of de novo synthesis of high-quality, gene-length DNA would allow for covert creation of sequences that might otherwise be subject to regulatory control or at least follow-up screening. Purchase and operation of these devices should be subject to the exact same expectations around appropriate customer- and sequence screening that centralized synthesis providers are subject to under the current or revised Guidance. Specifically, sale of these devices should be subject to export control license requirements and, domestically, companies should be restricted from selling the devices to private citizens or shipping to residential addresses. This consistency will be important in ensuring no easy obvious avenue for acquisition of otherwise controlled sequences by parties interested in misuse. This degree of regulatory control over privately owned devices is not new: scanners and some printers and copiers will recognize attempts at duplicating currency and refuse to do so.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 3-5

Next-generation benchtop DNA synthesizers could well pose similar biosecurity risks to those posed by mail-order synthetic DNA. Revised Guidance could, at minimum, signal its intention that such benchtop synthesizers should incorporate a screening mechanism. These need not be onboard the synthesizer itself. The synthesizers could be designed so that they must be connected to a screening server hosted by the manufacturer in order to operate.

In addition, manufacturers should undertake careful customer screening to ensure that purchasers of such equipment are legitimate scientific users. Finally, resale or transfer of equipment should require transfer of the synthesizer account to a new party who also undergoes screening. If the synthesizer must connect to a screening server, the account login process for such a server provides a ready point of intervention for actively tracking equipment and maintaining customer accounts.

Submitted on: 10/24/2020 9:07:31 PM

Agency Type: Other / Agency Other: Combined response from academic and company/business researchers

---



## 3-6

The biosecurity risks are technically the same whether the nucleic acid materials are produced via an external vendor or in-house on a benchtop synthesizer/assembler. Desktop engineering with sequence control at the local level must likewise be addressed in the Guidance, as it may become available in the near future.

The new Guidance needs to consider how much screening of materials to be synthesized or engineered locally is feasible and how technology can be used to attempt to prevent unauthorized misuse. A secure internet connection to a central risk screening authority is one potential option; alternatively, equivalent screening could be performed locally via a secured local system updated from a central authority (similar to how virus scanning software is regularly updated on our PCs.)

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 3-7

Widely distributed benchtop nucleic acid synthesizers capable of producing gene-length sequences (or oligos that can be used to assemble gene-length sequences) pose a grave biosecurity threat that is substantially different from the threats associated with a centralized DNA synthesis service. We strongly recommend that such synthesizers be regulated as critical infrastructure and not made broadly available for sale to end users because the risk of accidental or intentional misuse is too great. Please consider the following threat models and mitigations:

**Threat #1: sequence manipulation prior to synthesis.** Benchtop synthesizers must be connected to the internet in order to screen sequences according to the Guidance, especially since the databases of hazardous sequences must be continually updated as our knowledge of what is hazardous expands. However, as is painfully evident from the history of computing, every computer that has ever been connected to the internet at some point has had a vulnerability that could be exploited remotely, despite literally billions of dollars spent on securing these systems. Therefore, it is inevitable that any internet-connected DNA synthesizers will have windows of vulnerability in which they may be breached and remotely manipulated, silently. Average scientists certainly will not be able to fend off such attacks or even detect them. Moreover, it is the norm, not the exception, that scientists fail to perform timely software updates on their instrument controllers for fear of interrupting their functioning or losing access to deprecated features, greatly widening the windows of vulnerability of such instruments. A hacked instrument could be instructed to slip a malicious sequence into a synthesis run without knowledge of the scientist, who could then create a dangerous organism or virus inadvertently by inserting the synthesized DNA into a host. This type of attack will become increasingly practical as new types of high-fidelity DNA synthesizers streamline laboratory workflows, obviating the need for sequencing intermediate DNA products before performing an experiment.

**Mitigation #1:** ensure independent sequence-verification of synthesized products. Because a hacked benchtop synthesizer could silently insert hazardous sequences into the products, there must be an independent verification of the sequence of each synthesized product prior to its release to the end user. Any benchtop synthesizer absolutely must have an independent DNA sequencer onboard that it uses to verify what sequences it has actually synthesized. That sequencer must be controlled by a completely independent computer system from the synthesizer control system. Both control systems must be protected by a third independent system comprising a firewall that tightly controls incoming connections.

**Threat #2: “jailbreaking” of synthesizers or extraction of reagents.** With physical access to a benchtop DNA synthesizer, it is all but guaranteed that a motivated malicious actor could break any DRM or security controls on the instrument that prevent execution of synthesis of prohibited sequences. Furthermore, once the benchtop nucleic acid synthesizers and their reagents and processes become more robust, it will become practical for a malicious actor to simply crack open an instrument and/or reagent consumable in order to run the synthesis process manually or on other hardware.

**Mitigation #2:** strict access controls. The only mitigation for this threat is to ensure that the synthesizer and its consumables never fall into the hands of a malicious actor. This will be challenging to enforce if the instruments are broadly available for sale to end users, who may subsequently sell them to third parties without knowledge of the Government.

**Superior physical and network security of a centralized model.** In contrast to benchtop DNA synthesizers, a centralized DNA synthesis service can invest significantly in guarding physical and network access to the synthesizers and can ensure that the perimeter defences are kept up to date and vigilant. Furthermore, a centralized service can put into place processes for independently sequence-verifying every DNA product before shipping to a customer, taking advantage of sequencing economies of scale. Given the importance of securing these powerful instruments, the

Government should support small DNA synthesis businesses in their cybersecurity and physical security efforts around these machines.

Submitted on: 10/25/2020 5:05:30 AM

Agency Type: Company/Business / Agency Other:

---

## 3-8

Yes; in-house synthesis may allow malicious actors to completely bypass screening processes, even more so than the current ability to “venue shop” between providers. Encryption techniques must be developed to allow benchtop synthesizers to screen against a common database without creating intellectual property concerns. Two recent proposals on homomorphic encryption ([Titus et al., https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006454](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006454); [Esvelt, https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007286](https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007286)) propose methods by which sequences could be tested against databases while remaining encrypted throughout.

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 3-9

Yes. New generations of benchtop synthesizers are entering the market and should be included in the Guidance. How to do so is complex and, here, we give three (3) potential cases and provide pros and cons to each. Case 1: Benchtop synthesizers are subject to the same expectations for sequence and customer screening practices as a commercial entity. Advantages: (1) Clarity across the industry: All products are regulated the same way regardless of the production source or method. (2) Uniform screening ensures a consistent definition (subject to update over time) of what sequences are subject to regulatory control; this consistency does not incentivize circumvention of screening. Disadvantages: (1) Regulating inherently different technologies as the same can hamper innovation. Case 2: Benchtop synthesizers are subject to some of the expectations for sequence and customer screening practices as a commercial entity. Advantages: (1) Rules can be tailored to differences. Disadvantages: (1) Inconsistent guidance can lead to confusion: benchtop devices with 'onboard' sequence screening may just be required to look for potential homology above a certain threshold to controlled organisms while centralized providers are held to a 'best match' standard. This would result in sequences failing biosecurity screening at centralized providers but being accepted on a benchtop device and vice versa (for sequences not unique to controlled pathogens, which a benchtop device checking a blacklist would never know). (2) As benchtop device capabilities advance, the mismatch in expectations of these devices versus centralized providers may cease to be meaningful. (3) Distinctly less restrictive treatment of benchtop devices creates incentive for their use in gray-area scientific endeavors. Case 3: Benchtop synthesizers are subject to only institutional or internal review, and not the same expectations for sequence and customer screening practices as a commercial entity. Advantages: (1) Large organizations can leverage existing control systems: Many institutions have robust, existing systems that can be used as part of the risk assessment framework (e.g. IBCs). Disadvantages: (1) Lack of clarity across the industry: Actors who may not be able to pass sequence screening can bypass the rules to gain access. (2) Inconsistent Institutional Review Framework: Internal culture, history, and experience can impact decision making. IBCs, for example, may yield different results at different institutions. (3) Disincentivizes manufacturers from ensuring their devices are used responsibly. Assuming IBCs will provide oversight encourages device manufacturers to think of biosecurity as someone else's problem. (4) Incentivizes consumers of synthetic DNA who find biosecurity screening onerous to seek out these devices explicitly to avoid screening, whether or not they intend any misuse.

Submitted on: 10/25/2020 11:47:24 AM

Agency Type: NGO / Agency Other:

---

## 3-10

One major assumption of current regulation is that gene synthesis will primarily be handled by companies in a centralized process. Given the economics and convenience of gene synthesis using companies, companies might be responsible for synthesis of genes/oligonucleotides going forward rather than turning to a decentralized approach in which benchtop synthesis is performed within individual laboratories or university cores. However, this possibility is worth acknowledging. Regulation of these machines might need to occur on the software-level (eg. block synthesis of genes matching known pathogens, etc. in house and force reliance on any company to make the gene such that appropriate screening can be performed). This might also occur on the sales-level (preventing sales/exports of machines unless appropriate screening takes place).

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 4-1

**Technologies Subject to the Guidance - Question 4: As synthetic biology becomes an increasingly digital enterprise with large databases, digital tools, robotics, and artificial intelligence, what new risks are presented to providers and consumers of synthetic oligonucleotides?**

---

There are several considerations on this point. First, one potential issue is that sequences (DNA, RNA, protein) could be designed (either via mechanistic models or AI/ML), that are functional, yet have very little resemblance to any “known” sequence. That will make it very challenging to assess risk via the approaches above (best match, curated databases, predictive tools). However, there could be additional supplemental approaches - e.g. recognizing that there is no good match to any known sequence - which could prompt the screener to follow up with the end user/customer. Second, another issue is that the likelihoods (and perhaps even severity in terms of scale/quantity) in risk assessments made need to take into consideration of what these new technologies enable/facilitate.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 4-2

Cyberattacks can potentially be used to change a sequence being constructed to something different than what was screened. For example, a customer ordering a benign gene might instead be delivered the gene for a controlled toxin. This could be executed with either an attack on the screening system (substituting a benign sequence for a controlled order) or an attack on the order and synthesis workflow (substituting a controlled sequence for a benign order).

Such an attack could be used in multiple ways. For example, the attacker might wish to obtain controlled biological material that would otherwise be inaccessible to them. Alternately, the attacker might use control of synthesis to conduct a biological attack on a third party, by causing an unprepared laboratory to culture organisms producing a dangerous toxin.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---



## 4-3

With increasing access to annotated collections of biological components and automated design tools, actors interested in misuse of biology may be able to construct networks of otherwise 'harmless' sequences with emergent harmful properties, e.g. engineered gut flora that excrete immune modulators. Providers of synthetic DNA would need access to a new generation of biosecurity-related algorithms that could predict these harmful outcomes given a collection of biological functions, much as modern-day cybersecurity tools attempt to estimate risk by looking for the presence of collections of known components in binary executables.

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 4-4

While deliberate bioterror is always a concern, the increasing digitization of synthetic biology greatly increases the risk of "bio-error", e.g., inadvertent creation of dangerous pathogens that could occur from unknown or unknowable consequences of sampling portions of genome-space that nature has either not yet sampled or has learned to avoid for unknown reasons.

Cybersecurity ought to be covered e.g., to ensure that FASTA sequences could potentially be swapped out after biosecurity screening.

New advanced knowledge of potential biorisk sequences may become available leading to the dual use scenario that the information could also be used to create harm.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 4-5

This development means that cybersecurity vulnerabilities may rapidly become biological vulnerabilities. In my experience, commercial biotechnology firms have largely been motivated to follow cyber and physical security practices due to concerns about intellectual property rather than biosecurity. Biosecurity guidance must increasingly acknowledge digital containment in addition to physical containment. There are some potential reductions in risk from this as well; in contrast to the concerns related to benchtop synthesizers, the digitization of biology has led to some centralization of technology (for example, a smaller number of large sequencing providers who are able to benefit from economies of scale, rather than most labs doing sequencing in-house).

Submitted on: 10/25/2020 8:35:26 AM

Agency Type: NGO / Agency Other:

---

## 4-6

Risks include the mis-prediction and mis-characterization of certain oligonucleotides or resultant proteins thereof. The movement towards digital enterprise also goes beyond synthetic biology and encompasses biosciences at large as several labs work to build high-throughput, high-resolution datasets of protein variants at the same time. Other risks might include invasion of privacy of the general public if variants of known disease variants are studied and are easily identifiable from patient samples.

However the emergence of these tools are highly valuable for partnering with industry/academia to make screening more robust for variants of pathogenic proteins that require screening or regulation.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other:

## 5-1

**Technologies Subject to the Guidance - Question 5: If new risks are evident, how should these risks be addressed, keeping in mind the potential impacts on providers, customers, and scientific progress?**

---

As suggested in our previous responses, it behooves the Government and industry to establish a coordinated partnership and invest in programs to begin training the American workforce within an operational setting. Such a Government:private industry model should foster the early identification and mitigation of emerging biological risks.

Submitted on: 10/13/2020 9:34:09 AM

Agency Type: Company/Business / Agency Other: Non-profit Contract Research Organization

---

## 5-2

As stated immediately above, screening for sequences that are not good matches to any “known” public sequence should trigger subsequent follow-up screening. Note that this would require something like a best-match approach (i.e. screen all public sequences for matches) rather than just curated database or predictive tools approaches. In terms of the risk assessment, the assessors would just need to consider a variety of scenarios including those in which a bad actor has access to in-house or cloud/service capabilities that impact the likelihood and severity of the identified risks.

Submitted on: 10/23/2020 12:39:08 PM

Agency Type: Other / Agency Other:

---

## 5-3

The 2006 National Research Council's report *Globalisation, Biosecurity and the Future of Life Sciences* examines trends and objectives of research in the life sciences and converging fields such as materials science and nanotechnology, that may enable the development of a new generation of biological threats. The report notes that: "The growing concern regarding novel types of threat agents does not diminish the importance of naturally occurring threat agents – for example, the "classic" category A select agents – or "conventionally" genetically engineered pathogenic organisms. However, it does mandate the need to adopt a broader perspective in assessing the threat, focusing not on a narrow list of pathogens, but on a much wider spectrum that includes biologically active chemical agents. The potential threat spectrum is thus exceptionally broad and continuously evolving – in some ways predictably, in other ways unexpectedly. The viruses, microbes, and toxins listed as "select agents" and on which our biodefense research and development activities are so strongly focused today are just one aspect of this changing landscape of threats. Although some of them may be the most accessible or apparent threat agents to a potential attacker, particularly one lacking a high degree of technical expertise, this situation is likely to change as a result of the increasing globalization and international dispersion of the most cutting-edge aspects of life sciences research." [emphases added] The report concludes that a broad array of mutually reinforcing actions implemented in a manner that engages a wide variety of communities are required to successfully manage the threats that face society. The envisioned approach is described as a broad-based, intertwined network of steps – a web of protection – for reducing the likelihood that novel technologies may be used successfully for malevolent purposes. The adoption and promotion of a common culture of awareness of the value of formal international treaties and conventions, including the Biological and Toxin Weapons Convention (BTWC) and the Chemical Weapons Convention (CWC) within the global life science community is recognised as an essential element of the process of fostering an effective web of protection. (Source: Institute of Medicine and National Research Council. 2006. [Globalization, Biosecurity, and the Future of the Life Sciences](https://doi.org/10.17226/11567). Washington, DC: The National Academies Press. <https://doi.org/10.17226/11567>) In 2018, the US National Academies of Sciences, Engineering and Medicine published a report titled *Governance of Dual Use Research in the Life Sciences: Advancing Global Consensus on Research Oversight: Proceedings of a Workshop* which reiterated the importance of a comprehensive approach spanning the entire life science research cycle, in order to ensure that biosecurity and dual-use risks are identified and managed in a timely and effective manner. In this regard, the report drew attention to the need for maximizing the impact of two inter-related sets of activities, namely promoting engagement with biosecurity among different life science stakeholders, such as academic and research institutions, professional societies and trade associations, funding bodies, private companies, science publishers, and relevant government agencies and undertaking sustained biosecurity education efforts in the life sciences. (Source: National Academies of Sciences, Engineering, and Medicine. 2018. [Governance of Dual Use Research in the Life Sciences: Advancing Global Consensus on Research Oversight: Proceedings of a Workshop](https://doi.org/10.17226/25154). Washington, DC: The National Academies Press. <https://doi.org/10.17226/25154>) It is thus evident that the effective governance of biosecurity risks in the life sciences in the 21st century requires harmonised, globally distributed, and adaptive mechanisms for monitoring and assessing the implications of life sciences advances for maintaining the international norms of biological and chemical prohibition enshrined in the BTWC and CWC. Coordinated stakeholder engagement and systematic implementation of biosecurity education and training programmes are key prerequisites for the establishment and effective functioning of such mechanisms.

Submitted on: 10/23/2020 12:50:41 PM

Agency Type: NGO / Agency Other:

---



## 5-4

Synthesis providers should carry out routine cyber-threat evaluations, ideally including third-party red-teaming. Another important defense is extension of the guidance beyond a single point of failure in synthesis providers to all organizations that produce or handle synthetic organisms, as described above.

Submitted on: 10/24/2020 3:03:35 PM

Agency Type: Company/Business / Agency Other:

---

## 5-5

The cybersecurity community has addressed a similar challenge by formally identifying known vulnerabilities and making these publicly available. This enumeration allows researchers and security professionals to rank vulnerabilities by severity and to focus mitigation efforts on those capable of doing the most damage. These mitigation efforts often render specific vulnerabilities no longer relevant in terms of risk, thanks to coordinated efforts across a broad response community. A similar model would apply to biological systems as well - identifying potential routes to cause harm along with known biological systems that exploit these routes, e.g. binding of ribosomes to prevent protein expression, a route to harm that ricin and other ribosomal binding domain-containing toxins exploit. Ranking these vulnerabilities by severity and focusing investment on surveillance, diagnostic, and therapeutic approaches can mitigate their severity. As new sequences or networks of sequence components are identified that exploit known vulnerabilities in a given biological system, DNA synthesis providers could be alerted (in the same way that we alert network- and data center operators to new malware in the wild).

Submitted on: 10/24/2020 8:12:04 PM

Agency Type: Company/Business / Agency Other:

---

## 5-6

Increased guidance pertaining to the biosafety needs of massive, automated, high-throughput screening of combinatorially-modified organisms is needed to prevent inadvertent release of modified organisms with phenotypes that are dangerous to human/animal/plant/environment health.

The guidance should have some flexibility to address new biorisks that become evident as technology expands.

Submitted on: 10/24/2020 9:35:31 PM

Agency Type: Company/Business / Agency Other:

---

## 5-7

Such a database would need to effectively curate sources so that users can adequately assess conclusions made by others about proteins, and weights for how reliable conclusions/assessments in large databases are for certain proteins.

Appropriate anonymization would also be need to handled to protect the public's privacy. As far as I can tell, resources including cbioportal, gnomAD, and TCGA do a decent job of this.

Submitted on: 12/30/2020 7:19:50 PM

Agency Type: Academia / Agency Other: