HHS *Screening Framework Guidance for Providers and Users of Synthetic Oligonucleotides*, Summary of Updates in Response to Public Comments Received in 2020

In 2020, HHS began considering how changes in technologies during the past 10 years may necessitate changes in a critical biosecurity policy, the 2010 HHS *Screening Framework Guidance for Providers for Synthetic Double-Stranded DNA* (Guidance). To ensure that updates to this Guidance incorporate the viewpoints of experts in the academic and industrial sectors, HHS collected stakeholder responses to a 2020 Federal Register Notice titled *Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA* and undertook a deliberative process with its interagency colleagues to propose updates to the Guidance. In 2022, the revised guidance was published as a Federal Register Notice, titled *Screening Framework Guidance for Providers and Users of Synthetic Oligonucleotides* to solicit feedback on these proposed updates on the 2010 Guidance. Following is a summary of how public comments from 2020 elicited certain considerations and outlines the justification for the updates.

Scope of the Guidance

Most stakeholder responses agreed on extending the scope beyond the Federal Select Agent Program (FSAP) and Export Administration Regulations' Commerce Control List (CCL) agents. However, some commenters also mentioned that if the scope were expanded, there may also be negative impacts, including increased cost/burden of screening and potential negative effects on research and the bioeconomy. Responses centered on methods to define sequences of concern (SOCs) in a manner that enables consideration of sequences that may pose biosecurity risks, but that are not included in the FSAP and CCL lists. While some responses posed well-developed and comprehensive mechanisms to curate inclusion of sequences into a database of SOCs and assess whether the compilation of those sequences should pose restrictions on an order, there was not unanimity in whether this approach would be practical given the curation needs of databases, inadequate current infrastructure, and the immaturity of approaches based on predicting functionality. It was also noted that expanding beyond lists may make the screening easier, since it can be difficult to apply certain criteria in the Guidance that exclude housekeeping genes from being SOCs, even when they are from agents on the FSAP or CCL lists. As making the determination about these types of housekeeping genes is not always trivial, stakeholders report that it is the main source of the added cost of follow-up screening. This can require detailed knowledge of the function of sequences that would be included in the schema suggested for expanding beyond list-based approaches. Most respondents also thought that Guidance should be expanded to other providers in the supply chain, such as third-party sellers or transfers of SOCs beyond the initial purchaser.

As a result of these suggestions and concerns, the definition of SOCs in the Revised Guidance has been expanded, and now includes both sequences from agents in the FSAP or CCL lists and sequences that contribute to the pathogenicity or toxicity of agents not included in these lists. These changes were made to encompass the suggestions made by stakeholders and will aid synthetic oligonucleotide providers and customers in complying with best practices to ensure biosecurity.

In addition, as a result of these suggestions the Guidance now includes advice for verifying the legitimacy of recipients for transfers of either oligonucleotides containing SOCs or of oligonucleotide synthesis equipment, throughout the lifecycle of these materials.

Sequence Screening Methodology

The majority of respondents reported that the 200 base pairs (bp) window should be reduced to allow embedded sequences to be found. There was no clear consensus regarding the window size, but multiple responses indicate that screening window should be reduced to 40−50 bp. Some commenters find that 200 bp window is sufficient, and that reducing the size of the window should be subjected to cost/benefit analysis. Most respondents indicated that BLAST seems acceptable as a screening tool, but the lack of a definitive database of biothreat sequences was identified as a gap. Proteomic and bioinformatics approaches to curate a database of biothreats sequences and potentially establishing new methodologies other than BLAST were suggested. Regarding Best Match flagging, some stakeholders prefer to continue using Best Match with small modifications while others suggest moving to signature based detection or other approaches. An outcome-based approach was also suggested, where the government sets performance standards and then evaluates the screening methodology against them. Several responses indicated that, while predictive bioinformatic approaches are in development, they are likely not sufficiently mature on their own to identify SOCs. Respondents also indicated that the lack of a curated database of biothreat sequences outside the FSAP and CCL lists makes the development of these approaches unlikely. Several commenters supported the consideration of order batch size and indicated that the Guidance should include a discussion on the use of synthetic DNA fragments of a certain size for assembly into longer constructs and that providers should monitor Best Matches both a) across sequences within an order and b) across orders from the same customer.

In response to these suggestions, the Revised Guidance continues to recommend using Best Match for identifying SOCs, but also includes the consideration that Providers may also choose to use other screening approaches that they assess to be equivalent or superior to the Best Match approach or supplement it, including customized databases or approaches that evaluate the biological risk associated with non-select agents and toxins sequences or, for international orders, sequences not associated with items on the CCL. Also, the Revised Guidance acknowledges that

Providers may wish to consider developing solutions for determining which sequences from pathogens, regulated or unregulated, should not cause concern. The Revised Guidance also suggests that batch size should be considered in the sequence screening, to identify orders of small oligonucleotides that could be assembled into larger SOCs.

Biosecurity Measures

Many comments indicated that the maintenance and implementation of broader list-based approach(es) are now feasible and several curated SOC databases exist that have these functionalities. Several such databases were noted, and respondents listed their strengths and limitations. Other respondents noted that no single, recognized database exists. Respondents universally indicated that a curated database of sequences directly subject to regulatory control would be extremely valuable to Providers, and some respondents suggested that this database should be actively maintained in perpetuity by the U.S. Government – up to and including same-day updates coinciding with additions or removals of organisms on FSAP or CCL lists. Supplementing the Best Match approach with curated databases (beyond SAR/CCL) and predictive tools was suggested by some respondents, but the negative aspects of curated databases and predictive tools that may underestimate the hazard of sequences not in the curated databases or not predicted by tools was noted. Respondents also noted limitations in implementing broader list-based approaches. For example, the lack of operationally trained biosecurity experts was noted as a gap in broadly implementing these methodologies, and workforce training was mentioned as a potential area for U.S. Government investment.

In response to these suggestions and concerns, the Revised Guidance has been updated to include sequence screening recommendations that do not rely upon the use of curated databases of sequences that meet the definition of SOCs. However, the use of such databases is also not precluded, should they become more widely available. The U.S. Government encourages the development of such databases as screening tools that could be improved as additional data become available. Furthermore, in order to relieve some of the burden of sequence screening, the Revised Guidance includes the recommendation the recommendation that customers notify Providers if their oligonucleotide orders contain SOCs and provide proof of their legitimacy when placing such orders.

Customer Screening

Several respondents recommended consideration of methods to streamline customer screening. Pre-screening, white-listing, and the maintenance of a restricted list were suggested as potential responsibilities of the U.S. Government, and the risks and burdens associated with the white-listing were discussed in detail. Some respondents

supported pre-screening of the customer first, and others supported pre-screening of the sequences first. Some indicated that enhanced customer screening would require a registration program, and that whitelisting of customers may be helpful. Difficulty with tracking international orders was also noted. While most comments did not indicate that the Guidance poses an undue burden, the marginal cost of screening was noted to have increased as the overall cost of oligonucleotide synthesis has decreased. Many comments mentioned that manual review of false positive findings is the greatest cost of screening, since expert staff are expensive.

In order to relieve some of the burden of sequence screening, the Revised Guidance has been updated to include advice that customers notify Providers that their oligonucleotide orders contain SOCs and also provide proof of their legitimacy when placing such orders.

Minimizing Burden

According to respondents, implementing the Guidance is expensive, costing approximately $15/order, and this cost has remained flat while the cost of synthesis has decreased. Cost is reportedly driven by the need for a Ph.D. in Bioinformatics to interpret sequence hits. The burden of screening may also be increasing as the size of databases against which to match BLAST sequences is increasing. Providers indicated that they need new annotated data resources, tools, and approaches to keep biosecurity from becoming a leading component of the per-bp cost, given that oligonucleotide synthesis costs have consistently decreased.

Most respondents indicated that data retention is not a challenge, given ever-decreasing data storage costs. However, better guidance about what types of Customer and sequence data must be retained and what latency is acceptable for retrieval of this information would be helpful. Highly latent data storage mechanisms are much less costly, especially given customer screening considerations that are also included in the Guidance. The eight-year timeframe for data storage may be a burden for start-up companies, and mitigation processes are necessary in case they are no longer in business after eight years.

Some respondents asked the U.S. Government to provide standardized screening methodologies and a centralized database for screening both sequences and customers. These requests included clear definitions of the unit of control and of the type of information needed for screening customers. They also included requests for databases of biothreats sequences and a centrally located white-listed or restricted list of customers. There was also a suggestion that the US Government should operate an Application Programming Interface (API) for these screening concerns that has a latency for queries of two hours or less. Respondents also expressed that expanding the scope of the Guidance would not create additional burden if the U.S. Government took some of these steps, but others were concerned about potentially increased costs

associated with expanding the scope of the Guidance. Support for a cost/benefit approach was voiced to ensure that any additional burden to Providers (and customers) can be adequately justified. There were mixed comments regarding whether liability is a concern, with one comment mentioning that the Guidance is perceived as protecting Providers from liability if it is followed.

In response to these suggestions and concerns, the Revised Guidance includes sequence screening recommendations that do not rely upon the use of curated databases of sequences that meet the definition of SOCs. However, the use of such databases is also not precluded, should they become available. The U.S. Government encourages the development of such databases as screening tools that could be improved as additional data become available. Furthermore, in order to relieve some of the burden of sequence screening, the Guidance has been updated to include the recommendation that customers notify Providers if their oligonucleotide orders contain SOCs and provide proof of their legitimacy when placing such orders.

Also, in response to these concerns, the Guidance includes specific suggestions for who should store records of screening methodologies, hits identifying SOCs in orders, the proof of Principal User or End User legitimacy, and records of transfer of oligonucleotides containing SOCs to new End Users beyond the original customers or Third-party Vendors. There is no requirement for latency in the retrieval of this information, so concerns about high latency versus low latency storage costs may be alleviated by the Revised Guidance.

Technologies Subject to the Guidance

Due especially to the ease of conversion between single-stranded (ss) DNA, double-stranded (ds) DNA, ssRNA, and dsRNA; that positive sense ssRNA viral genomes can be transferred directly into cells to produce viruses; and that rescue platforms exist for negative sense ssRNA viruses using modern methodologies, all respondents suggested screening each type of synthetic oligonucleotide orders, not just dsDNA.

Some respondents recommended that the Guidance apply to the entire synthetic biology supply chain, not just to the Providers of synthetic DNA or other oligonucleotides. Also, some responses indicated that benchtop DNA synthesizers pose a serious biosecurity threat. A molecular biology-based biorisk approach was described by several respondents, some suggesting that this approach would benefit from a centralized U.S. Government sequence screening database.

In response to these comments, the Scope of the Guidance has been expanded beyond dsDNA to also include single and double-stranded forms of both RNA and DNA. Also, the Guidance has been expanded to include advice for Providers, Third-party Vendors, Principal Users, and End Users as well as Manufacturers of bench-top oligonucleotide synthesis equipment.

Also, in response to these suggestions, the Guidance has been expanded to include a section titled *Periodic Review, Evaluation, and Improvement of this Guidance*. This section of the Revised Guidance includes recommendations that methodologies should be developed to use predictive bioinformatics algorithms screen sequences that are not Best Matches to any known sequences – especially if no explanation is provided by the Customer – to determine whether they could produce proteins that are structurally or functionally identical to SOCs, and also that sequences that can confer medical countermeasure evasion to pathogens or toxins should be identified for potential future inclusion in the SOC definition.