DNA SYNTHESIS AND BIOSECURITY:
Lessons Learned and Options for the Future

Sarah R. Carter, Ph.D. and Robert M. Friedman, Ph.D., *J. Craig Venter Institute, La Jolla, California*

This project was funded by the Alfred P. Sloan Foundation.

October 2015

J. Craig Venter®
INSTITUTE

The views and opinions expressed in this report are those of the authors and not necessarily those of the institution at which they work, the participants in the workshop discussed in the report, or the organization that funded the study. The authors assume full responsibility for the report and the accuracy of its contents.

Table of Contents

Abstract

Synthetic biology promises great scientific advances, but it also has the potential to pose unique biosecurity threats. It now is easier than ever to synthesize very long pieces of DNA from chemicals, potentially enabling a bioterrorist to build a toxin gene or an entire pathogenic virus. To guard against this possibility, the Department of Health and Human Services released its "Screening Framework Guidance for Providers of Synthetic Double-stranded DNA" in 2010, which called on DNA providers to screen both customers and the DNA sequences ordered by those customers for potential biosecurity concerns. In this report, we evaluate how well the Guidance has been implemented by DNA providers and consider changes that could be made to the Guidance so that it can keep pace with anticipated changes in the DNA synthesis industry. While the Guidance has worked reasonably well over the past five years, we identify two options that policy makers could pursue to strengthen the Guidance for the next five years: 1) require federal grantees and contractors to purchase double-stranded DNA only from companies that comply with the Guidance, and 2) provide a curated database of "sequences of concern" for DNA providers to use for screening. We also consider ways in which the Guidance could be expanded to address short, single-stranded DNA ("oligos") and benchtop synthesizers capable of making double-stranded DNA

Introduction

The methods of synthetic biology have become a mainstay of biological research, but along with these scientific advances have come potential new biosecurity threats. The heart of the issue is that DNA now can be synthesized rapidly and inexpensively from chemicals to construct, for example, toxin genes or entire viral genomes that could be used for nefarious purposes. These techniques potentially could allow a bioterrorist to build a virus from knowledge of its DNA (or other nucleic acid) sequence without the need to obtain a physical sample of the virus itself.[1]

Early discussions of this potential biosecurity threat included ideas for how the U.S. government and others could address it. In 2006, the National Science Advisory Board for Biosecurity (NSABB) recommended that the U.S. Government issue screening guidelines for orders of double-stranded DNA (dsDNA), update the Select Agent regulations, and include synthetic DNA in guidelines for laboratory biosafety (NSABB, 2006).[2] In our 2007 report, "Synthetic Genomics: Options for Governance," we evaluated a variety of options, including policies for commercial DNA synthesis firms, policies to monitor or control equipment or reagents, and policies for users of synthesized DNA and their institutions (Garfinkel, et al., 2007). Although our report did not make recommendations, we found that the screening of dsDNA orders and storage of information about customers and their orders by DNA synthesis companies were among the options that would "provide the greatest benefits at the lowest costs and burdens."

Ultimately, in 2010, the U.S. Department of Health and Human Services (HHS) issued the "Screening Framework Guidance for Providers of Synthetic Double-stranded DNA" (Guidance), which calls on providers of dsDNA to screen both customers and ordered DNA sequences (HHS, 2010). (See Box A.) Now, five years later, we have undertaken this review of the Guidance with two goals in mind: 1) to evaluate how well the Guidance has worked during its first five years, and 2) more importantly, to consider whether changes in the Guidance might be needed to keep pace with anticipated developments in the field of DNA synthesis over the next five years.

In issuing its Guidance, HHS allowed quite a bit of flexibility in how biosecurity screening would be implemented in the U.S. The Guidance has been reasonably successful with a large majority of the industry in voluntary compliance. Below we discuss some of the ways in which companies have implemented

The heart of the issue is that DNA now can be synthesized rapidly and inexpensively from chemicals to construct, for example, toxin genes or entire viral genomes that could be used for nefarious purposes.

1 To our knowledge, a comprehensive risk assessment never has been done for this potential threat. At present, the construction of toxin genes, pathogenic pathways, and many viral genomes is feasible for a well-trained scientist in a reasonably well-equipped laboratory. In 2007, we concluded that although many viruses were easier to obtain in nature than construct in a lab, there were a few exceptions (Garfinkel, et al., 2007). Since that time, laboratory techniques have greatly improved, making construction of viral genomes easier. Construction of a bacterial genome, in contrast, remains an overwhelming task and has been accomplished by only a single group (Gibson, et al., 2010).

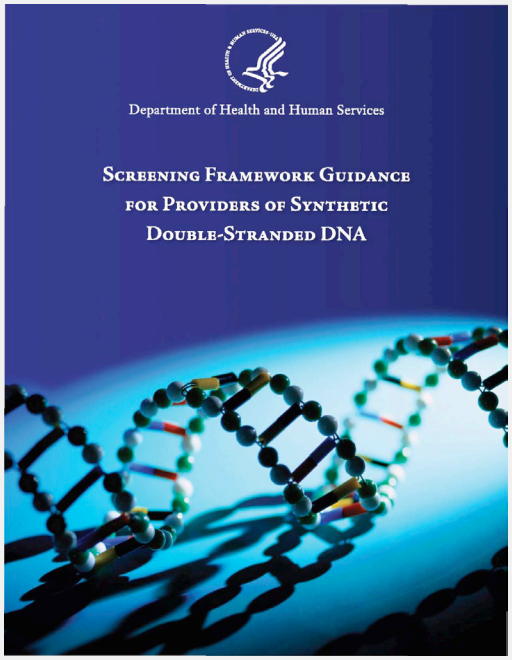
2 Since that time, the National Institutes of Health (NIH) Guidelines that govern laboratory biosafety have been updated, and now are titled "NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules" (NIH, 2013).

The Guidance has been reasonably successful with a large majority of the industry in voluntary compliance.

Box A: HHS Screening Framework Guidance for Providers of Synthetic Double-stranded DNA

Briefly, the Guidance to providers of synthetic double-stranded DNA states:

"Providers should establish a comprehensive and integrated screening framework that includes both customer screening and sequence screening, as well as follow-up screening when customer and/or sequence screening raises a concern.



- *Customer Screening* – The purpose of customer screening is to establish the legitimacy of customers ordering synthetic dsDNA sequences. Providers should develop customer screening mechanisms to verify the legitimacy of a customer if the customer is an organization or confirm customer identity if the customer is an individual, to identify potential 'red flags,' and to conform to U.S. trade restrictions and export control regulations.
- *Sequence Screening* – The purpose of sequence screening is to identify when "sequences of concern" are ordered. Identification of a "sequence of concern" does not necessarily imply that the order itself is of concern. Rather, when a "sequence of concern" is ordered, further follow-up procedures should be used to determine if filling the order would raise concern. Sequence screening is recommended for all dsDNA orders.
- *Follow-up Screening* – The purpose of follow-up screening is to verify the legitimacy of customers both at the level of the customer and the principal user, to confirm that customers and principal users placing an order are acting within their authority, and to verify the legitimacy of the end-use."

(Italics original.)

the Guidance and the reasons why a few DNA providers have implemented them only partially or not at all. In addition to describing the current status of biosecurity screening in the U.S., we also outline changes in the industry that may impact the future of the Guidance and options that policy makers can pursue to anticipate these changes.

Over the course of this project, we have had numerous conversations with industry representatives, stakeholders, and policy makers. In addition to these discussions, we held a workshop on April 28, 2015, in Washington, D.C. The workshop was attended by experts from industry and government as well as outside biosecurity scholars. During that meeting, we learned about the biosecurity practices that many in the industry have adopted and the administrative burdens associated with voluntary compliance with the Guidance. These burdens are primarily due to the professional staff time and costs devoted to screening and follow-up activities. We also discussed the challenges that policy makers face in improving adherence to the Guidance and in addressing biosecurity threats related to synthetic biology that are not covered, including the potential use of short, single-stranded DNA ("oligos") and benchtop DNA synthesizers to make dsDNA. Because the DNA synthesis industry is rapidly changing, we used the gathered expertise to better understand where the industry is going and how biosecurity challenges will evolve and may be met.

Although the development of the project and the writing of this report were heavily dependent on and influenced by the input of others (see Acknowledgments), the conclusions and options written in this report are our own. No consensus was sought or obtained.

In addition to describing the current status of biosecurity screening in the U.S., we also outline changes in the industry that may impact the future of the Guidance and options that policy makers can pursue to anticipate these changes.

HHS Guidance: Original Choices and their Implications

When HHS chose to issue the Guidance to providers of dsDNA, it linked the success of biosecurity practices with the US. gene synthesis industry.

The choices that the U.S. government made when first developing the Guidance have had important implications for its success and limitations. When HHS chose to issue the Guidance to providers of dsDNA (rather than, for example, putting some onus on institutions or other users), it linked the success of biosecurity practices with the U.S. gene synthesis industry. This partnership has been reasonably successful to date because established companies are highly motivated to prevent any biosecurity mishaps that could implicate their firms or their industry. In our conversations with industry representatives, we repeatedly heard their concern that any biosecurity lapse on their part could result in a public outcry, legal liability, and/or government action that would severely restrict not only an individual company but the industry as a whole with national and international significance. Indeed, two different gene synthesis industry groups—the International Gene Synthesis Consortium (IGSC) and the International Association for Synthetic Biology (IASB)—had developed biosecurity codes of conduct with screening procedures before the HHS Guidance was finalized (IASB, 2009; IGSC, 2009).[3]

Because compliance with the Guidance is voluntary and no reporting is required, biosecurity practices are not very transparent or available to the public or to the government. While the IGSC and others are willing to share some information about their procedures as described below, details of how any one company has reviewed or will review any specific order remain unclear. The Guidance frequently suggests that providers contact the FBI or the U.S. Department of Commerce for additional advice. However, because confidentiality often is critical to purchasers of DNA for reasons of commercial competitiveness, providers of synthetic DNA take customers' confidentiality concerns very seriously.

The Guidance built on existing regulatory frameworks surrounding the U.S. Select Agent Program and Commerce Control List and attempted to address "biosecurity concerns associated with the potential misuse of [synthetic dsDNA] to bypass existing regulatory controls." This narrow focus led to an emphasis, in the Guidance itself and in subsequent discussions, on the species listed on the Select Agent list and the Commerce Control List and their associated DNA sequences.[4] However, this approach has a great deal of ambiguity that still has not been resolved; because the Select Agent list and similar lists are composed of whole organisms, it remains

3 Although these codes follow the spirit of the HHS Guidance, especially with customer and DNA sequence screening procedures, as described below, they are adapted so that international companies also can adhere (i.e. the HHS Guidance recommends reporting to the U.S. Federal Bureau of Investigation, which is not realistic for non-U.S.-based companies).

4 There has been ongoing discussion about biosecurity concerns that can arise even outside the context of Select Agents and related lists, including not only other sources of pathogenicity but also more complicated biosecurity threats (for example, production of illicit drugs using engineered organisms; DeLoache, et al., 2015). However, because the DNA sequences relevant to these broader threats remain difficult to specify, both the HHS Guidance and DNA providers have primarily focused on the narrower set of sequences.

unclear whether and how DNA sequences from those organisms should be considered for DNA sequence screening. HHS addressed this issue in the Guidance in multiple places. The Guidance states that "sequences of concern" are dsDNA sequences "derived from or encoding" Select Agents and Toxins (p. 8), but it does not define what "derived from" means. In a section on Technical Goals and Recommendations, the Guidance elaborates that screening should identify sequences that are "unique" to Select Agents and Toxins and that "house-keeping genes" that maintain normal cellular physiology should be excluded (p. 9), but again, the Guidance does not identify or suggest how to determine if a sequence encodes a "house-keeping gene." The Guidance goes on to recommend that sequences that are a match to strains that are closely related to Select Agents but that have been specifically exempted from controls (such as attenuated vaccine strains) also should be considered "hits" (p. 11), implying that some measure of homology (i.e., sequence similarity) could be used.

HHS clearly understood that there was some ambiguity about which DNA sequences should be considered a biosecurity threat. The Guidance emphasizes in multiple places that best practices may be developed over time, both for screening algorithms and for determining what should be considered a sequence of concern, and that the Guidance itself may evolve. Over the past five years, this central uncertainty has remained a challenge for DNA providers, and we discuss it in more detail below.

The Guidance emphasizes in multiple places that best practices may be developed over time and that the Guidance itself may evolve.

HHS Guidance: Current Implementation by dsDNA Providers

The International Gene Synthesis Consortium (IGSC) is comprised of seven companies that as a group provide what they estimate to be about 80% of commercial gene-length synthetic DNA. The IGSC member companies have developed their own biosecurity practices that are in line with the HHS Guidance.

The International Gene Synthesis Consortium

The International Gene Synthesis Consortium (IGSC) is comprised of seven companies[5] that as a group provide what they estimate to be about 80% of commercial gene-length synthetic DNA (Graf, 2013; personal communications). The IGSC member companies have developed their own biosecurity practices that are in line with the HHS Guidance and that are described in their "Harmonized Screening Protocol" (IGSC, 2009). This Protocol describes the overall goals and procedures; each company implements the screening in the way that is best suited to its operations.

Each order for dsDNA that comes into an IGSC company undergoes both customer and bioinformatic sequence screening. The initial customer screening includes using software such as BridgerInsight from LexisNexis, which checks the name of the customer for matches against U.S. government watch lists for terrorists, those engaged in the trafficking of weapons of mass destruction, those debarred by HHS or the State Department, and others.[6] If the customer is flagged as a match to one of these lists (a rare occurrence), then the company will not fill the order.

Companies also check for an institutional affiliation and neither sell DNA to unaffiliated individuals nor ship to P.O. boxes. There is some ambiguity in what constitutes a legitimate research institution. For example, while companies do not fill orders for unaffiliated individuals, a well-established community lab (such as Genspace in New York; http://genspace.org/blog/) may qualify depending on who is making the determination and according to what standards.

In addition to the initial customer screening, the DNA sequence that is ordered goes through a two-step screening procedure. First, it is screened against GenBank (a database containing all publicly available DNA sequences, both pathogenic and non-pathogenic) to determine the best matches.[7] The top matches then are compared to a database of DNA sequences associated with pathogenic species.[8] Development of this "Regulated Pathogen Database" has been a major undertaking by the IGSC over the past several years, and it includes whole genomes of pathogenic species (including those on the

5 These companies are DNA2.0, Genscript, Gen9, Integrated DNA Technologies (IDT), Origene, SGI-DNA, and Thermo Fisher Scientific. Disclaimer: the authors of this report are with the J. Craig Venter Institute, which holds stock in SGI-DNA.

6 These lists include those administered by the U.S. Department of the Treasury's Office of Foreign Assets Control (Specially Designated Nationals (SDN), Non-SDN Entity List, and Sanctioned Countries), the list of debarred parties administered by the U.S. Department of State, and a list maintained by the Bureau of Industry and Security of the U.S. Department of Commerce. The consolidated list can be found at http://export.gov/ecr/eg_main_023148.asp.

7 The IGSC companies use a 200-base-pair sliding window on the ordered gene. To account for possible codon optimization or obfuscation, each 200-base-pair DNA sequence is translated to its six possible amino-acid sequences (based on three possible open-reading frames, both forward and backward). This six-frame translation is then screened against the amino-acid sequences in GenBank.

Select Agent list, the Commerce Control List, the Australia Group list, and the European Union's Council Regulation 428/2009)[9] as well as subspecies and variants of pathogenic species, toxin genes, and other DNA sequences that may be associated with pathogenicity. IGSC updates this database periodically.

The IGSC companies have adopted a common "red-yellow-green" approach for their bioinformatic screen. Ordered sequences that are not a match to a sequence in the Regulated Pathogen Database are considered "green," and the order may be filled without further complication. If any of the best matches for an ordered DNA sequence are a match to a sequence in the pathogen database (i.e., the sequence has approximately 80% homology or greater; see footnote 8), the sequence is considered a "hit." A bioinformatician then evaluates the hit to determine how the company should proceed. If the hit is not the best match, it still could be deemed "green," and the order filled without further follow-up. A "yellow" hit may have high homology (i.e., sequence similarity) to a pathogenic species but is not thought to contribute to pathogenicity (e.g., a "house-keeping" gene in a bacterial pathogen). A sequence is considered "red" when it is unambiguously linked to pathogenicity or toxicity. Although the IGSC itself gives guidance (and is working to improve its guidance) as to which sequences should be considered red, yellow, or green, this determination is made by the individual bioinformaticians at each company.

For each yellow or red hit, the company follows up with the customer to determine if the customer has a legitimate reason to order genes or gene fragments from pathogenic species. This follow-up could include direct email and/or phone correspondence, internet searches to determine if the customer is a researcher in a relevant field, requests for grant or funding information and/or sign-off on the order from others at the researcher's institution. If a sequence is unambiguously pathogenic (a "red" hit), the DNA provider will perform follow-up screening and will ensure that the customer obtains an export license, if needed. If the researcher cannot be verified or has no legitimate reason to order that DNA, then that order may be reported to the FBI. Such a situation has occurred only a handful of times since the establishment of the HHS Guidance. More often, the customer will withdraw or amend the order before the follow-up screening efforts of the DNA provider proceed to the point of FBI notification. The IGSC has implemented a suspicious-order communication mechanism so that a company can notify other IGSC members in such a case so that the order will be flagged if it comes to another company. This system is to foster communication and is not intended to replace law enforcement.

IGSC companies report that much of the administrative burden associated with following the HHS Guidance is due to the follow-up that is required by professional-level staff after red and yellow hits on the bioinformatic screen. Approximately 5% of orders are hits,

The IGSC companies have adopted a common "red-yellow-green" approach for their bioinformatic screen. For each yellow or red hit, the company follows up with the customer to determine if the customer has a legitimate reason to order genes or gene fragments from pathogenic species.

8 Although the HHS Guidance recommends that DNA providers evaluate the "Best Match" to see if it is a sequence of concern, the IGSC companies evaluate a long list of the best matches. At least one company screens the top 800 matches for possible homology to pathogens. It is estimated that using a long list of matches effectively screens sequences that have at least 80% homology with a pathogenic species, which is used as an informal goal among IGSC companies.

9 These are more organisms than are required by the Guidance, which lists only the Select Agent list and the Commerce Control List.

with fewer than 1% determined to be red hits. Yellow hits often require approximately 60-90 minutes to resolve, and each red requires several hours. (See Box B for an estimate of costs for bioinformatic screening.) It should be noted the IGSC companies spend significant time following up on yellow hits (sequences that are not directly pathogenic). As mentioned above, HHS sought to minimize time spent by companies on such sequences, stating in the Guidance that "house-keeping genes" should not be considered as hits even as the Guidance leaves "house-keeping genes" undefined. Further complicating the issue is that companies may request that an international customer seek an export license for any sequence related to a Select Agent, even if a bioinformatician may consider it to be a non-pathogenic, yellow sequence.

Approximately 5% of orders are hits, with fewer than 1% determined to be red hits. Yellow hits often require approximately 60-90 minutes to resolve, and each red requires several hours.

Box B: Estimate of Time Spent and Costs for Bioinformatic Screening

The numbers below are estimates based on data collected from IGSC member companies but do not represent any single company's orders or costs. Green, yellow, and red sequences are described in the text.

Time for Screening

Type	% of orders of this type	Bioinformatics review time	Customer follow-up	Cost, assuming labor @ $150/hour
"Green"	95%	0.5 min	0 min	$1.25
"Yellow"	4.3%	4.5 min	79 min	$209
"Red"	0.7%	7.5 min	232 min	$598

For any given order, a company can expect to spend, on average, $14.35 on bioinformatic screening and the necessary follow-up with the customer (based on an average of costs weighted by the percentage likelihood of green, yellow, and red sequences). Because genes often cost on the order of $500–$1,000, this screening plus follow-up represents approximately 1.5-3% of total costs. As the price of gene synthesis goes down, this percentage will increase.

Of the total time spent screening orders, approximately 13% is devoted to bioinformatics-review time to determine whether the sequence is red or green or the more ambiguous yellow. The remaining 87% of time is devoted to customer follow-up for the red and yellow sequences. Close to 60% of the total screening time is devoted to customer follow-up for orders that are unlikely to be able to cause harm (i.e. yellow sequences). A more selective definition of "sequences of concern" might lower screening costs by half or more. (See discussion on a database of sequences of concern on page 10.)

Other dsDNA Providers

Outside of the IGSC, there are a variety of different practices among dsDNA providers, some of which follow the HHS Guidance and some of which do not.[10] Several companies adhere to the "Code of Conduct for Best Practices in Gene Synthesis," developed by the IASB (IASB, 2009). This code takes a similar approach to that taken by the IGSC and describes procedures that follow the HHS Guidance, with customer screening followed by sequence screening that must be evaluated by trained bioinformaticians.[11] When an ordered sequence is determined to be a hit, the companies follow up with the customers to make sure that they are legitimate users of that DNA. Other small dsDNA providers, both commercial and non-commercial, voluntarily comply with the Guidance using procedures developed in-house.

We heard from at least two companies that rely primarily on customer screening and only screen DNA sequences from unknown customers. Once a customer is trusted (i.e., one from a recognized, legitimate institution, often with a prior relationship with the DNA provider), then the order will be filled without undergoing sequence screening. Although these companies believe that these procedures represent an appropriate level of due diligence, they do not adhere to the Guidance.

A major consideration for smaller companies in determining whether to fully follow the Guidance is the administrative burden associated with full compliance. Some level of customer screening, including screening against terrorist watch lists and other lists, is required by law for all U.S. companies.[12] Commercial software such as LexisNexis' BridgerInsight, mentioned above, has been developed to meet that need. However, further customer screening to determine if a researcher is with a "legitimate" institution requires some research and can be ambiguous. DNA sequence screening requires time and specialized knowledge both to establish a bioinformatic screening procedure and in an ongoing way to evaluate each order, with additional time spent following up on hits generated from the screening. (See Box B.) It has been difficult, particularly for smaller companies, to implement and maintain a bioinformatic screening procedure, especially given the equivocal criteria.

A major consideration for smaller companies in determining whether to fully follow the Guidance is the administrative burden associated with full compliance.

10 TIn addition to in-depth conversations with non-IGSC members (including a few who participated in our workshop), we reached out to 22 non-IGSC member companies through email to gauge knowledge of and adherence to the HHS Guidance. Although not all of them responded, several gave their perspectives.

11 The IASB companies screen each ordered sequence against GenBank, and the top matches are evaluated to determine if they are associated with pathogenic or toxic species. As with the IGSC protocol, there is flexibility for each company, including international firms, to implement the procedures in a way that best meets its needs.

12 A summary of required customer screening can be found at http://export.gov/regulation/eg_main_018219.asp.

Biosecurity Screening Tools to Support Voluntary Compliance with the HHS Guidance

Although some companies have developed their own procedures to follow the HHS Guidance, additional tools for dsDNA providers are becoming available. In an effort to expand its membership to the point that its screening procedures become the de facto industry standard, the IGSC incorporated as a non-profit organization, which was announced on April 28, 2015, the same day as our workshop (IGSC, 2015). dsDNA providers, including both commercial and non-commercial (e.g., academic) enterprises, will be able to join the organization and have access to its Regulated Pathogen Database and decision support tools. Because access to such a database would make screening easier for smaller companies, this development may improve biosecurity practices in the industry. At the time of writing, a number of commercial and non-commercial entities have applied for membership in IGSC.

Non-commercial dsDNA providers also have developed screening procedures and tools to follow the HHS Guidance. The U.S. Department of Energy's Joint Genome Institute provides dsDNA through its DNA Synthesis Science program, and has developed a protocol, including software, for biosecurity screening (Simirenko & Hillson, 2015). By the end of 2015, they plan to publish their process and experience to date and to make this software available to other credentialed researchers and dsDNA providers. This software represents an additional resource for smaller companies and other dsDNA providers who may otherwise find DNA sequence screening to be too burdensome.

The Changing Industry as a Key Challenge for the Future of the HHS Guidance

How the gene synthesis industry changes over the next five years will be critical to the future effectiveness of the HHS Guidance. The Guidance has done reasonably well over the past five years in part because the industry has evolved as was expected: remaining largely based in the U.S. and Europe and consolidated into bigger companies, and with dsDNA orders remaining relatively expensive ($100s to $1,000s of dollars per order) and slow to fill (more than a week). All of those market characteristics may be changing in ways that will make it more difficult for the Guidance to remain effective.

Over the past five years, dsDNA synthesis has become steadily cheaper, with costs falling from about $0.70 per base pair to between $0.10 to $0.30 per base pair (Carlson, 2014; see Figure 1). However, while the costs for synthesis have fallen, the administrative costs of biosecurity screening have not significantly changed, which means that biosecurity screening is an increasingly significant cost for dsDNA providers. Because the staff time required already is a key reason that smaller companies may not utilize sequence screening, this trend will make following the Guidance increasingly difficult. Furthermore, the computational resources required for sequence screening, which are cited as a challenge for some smaller dsDNA providers, will increase over time as the number of non-redundant DNA sequences in GenBank continues to increase rapidly. (Note that costs associated with these computational resources are not included in Box B.)

There is also the potential for disruptive technologies to be introduced into the marketplace as soon as this year that will dramatically decrease the costs for gene synthesis (by an order of magnitude or more) and the time required to fill the order.[13] When such technologies become available and widespread, the Guidance cannot be implemented in the same manner that it is today.

In addition to declining costs, the Guidance will be challenged by the increasingly international nature of the gene synthesis industry. All seven of the IGSC member companies have headquarters in the U.S. While IGSC member companies currently account for an estimated 80% of the global gene synthesis industry, international players, particularly Chinese companies, are rapidly increasing their share of the market. There is some interest among Chinese companies in implementing biosecurity safeguards, and some report that they have already,[14] but there is no consensus on how the industry should proceed. As the costs for DNA synthesis decline and screening procedures are proportionally more of a

How the gene synthesis industry changes over the next five years will be critical to the future effectiveness of the HHS Guidance.

13 Twist Biosciences (http://www.twistbioscience.com/) and Cambrian Genomics (http://www.sfgate.com/business/article/Controversial-DNA-startup-wants-to-let-customers-5992426.php) are two companies working to dramatically reduce the price of DNA synthesis.

14 Meetings were held in Shanghai in August, 2012, and in Hong Kong in March, 2013, that were well-attended by Chinese companies. At those meetings, there was interest among some of the participants in developing a Chinese code of conduct similar to the IGSC or IASB protocols (ICLS, 2013). However, there has been no discernable progress since that time.

burden, U.S. companies are increasingly concerned about competing with international companies that might not screen orders for biosecurity purposes.

Decentralization of DNA synthesis is another factor that may complicate the future of the Guidance. Commercially available chemical kits have simplified the assembly of gene-length dsDNA from short, single-stranded DNA ("oligos"). The recent introduction of benchtop DNA synthesizers capable of making dsDNA also may shift some fraction of the market to in-house assembly. In principle, absent new guidance, these lab-bench options may allow a potential bioterrorist to evade biosecurity practices more easily. Neither oligos nor benchtop DNA synthesizers, discussed in more detail below, are addressed by the current Guidance.

Again, the Guidance is working reasonably well today. The challenge it faces is ensuring that it remains as relevant into the future. All of the factors mentioned above—declining dsDNA synthesis costs, greater international participation in the industry, potential decentralization of dsDNA production—will make it more difficult for DNA providers to adhere to the Guidance and more difficult for policy makers to maintain a consistent level of biosecurity screening . Below, we discuss options that policy makers could pursue to improve voluntary compliance with the Guidance and to expand the Guidance to address oligos and benchtop DNA synthesizers.

Declining dsDNA synthesis costs, greater international participation in the industry, and potential decentralization of dsDNA production will make it more difficult for DNA providers to adhere to the Guidance and more difficult for policy makers to maintain a consistent level of biosecurity screening.

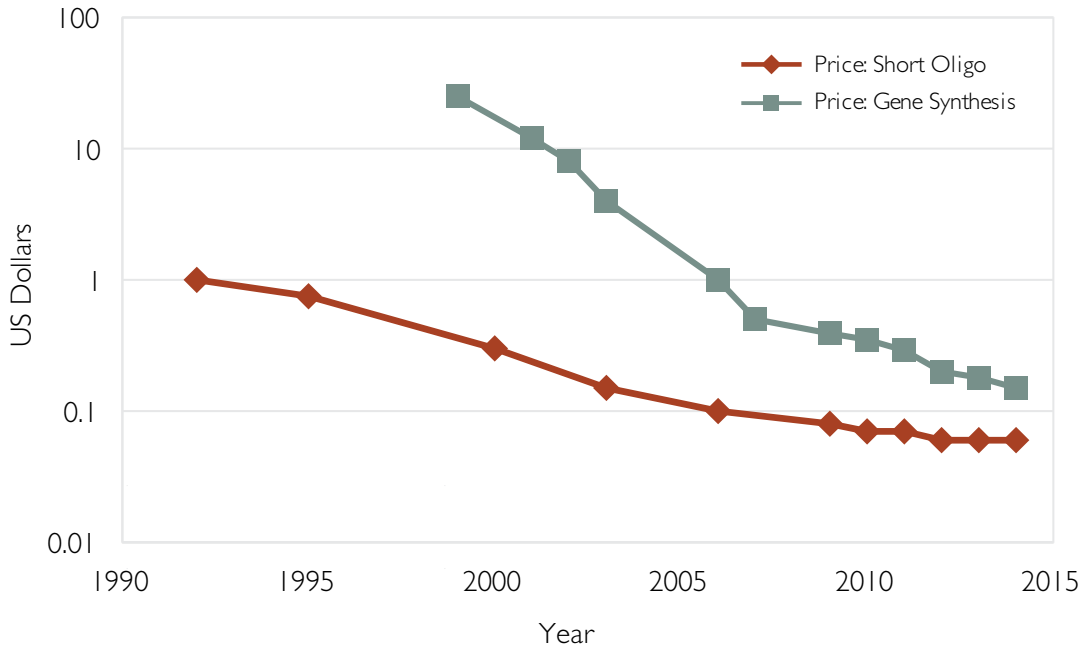


Figure 1: Reproduced from Carlson, 2014

Improving Voluntary Compliance with the HHS Guidance

Although most U.S.- and E.U.-based DNA providers (the IGSC members plus others) follow the recommendations of the HHS Guidance, there are many providers that do not. We spoke with at least two companies that rely on the trust developed with their customers and only rarely screen DNA sequences. It is likely that other commercial providers in the U.S. perform only the legally required minimal customer screening using government watch lists. Outside the U.S. and Europe, there may be even fewer companies practicing biosecurity screening procedures.

As discussed above, a major reason that smaller companies choose not to follow the Guidance is the administrative and commercial burden associated with compliance. Currently, the IGSC reports that DNA sequence screening requires significant time and expertise in determining matches with pathogenic sequences, and then more time to follow-up with customers to determine if they have a legitimate use for the ordered DNA. (See Box B.) Furthermore, each company's procedures depend on the judgment of individual bioinformaticians within a common decision framework that still may lead to different determinations about the same sequence.

We have identified two options that the U.S. government could consider to improve voluntary compliance with the Guidance across the industry and reduce the administrative and commercial burden imposed by biosecurity practices:

1. **Through contractual agreements, require federal grantees and contractors to purchase dsDNA only from companies that comply with the Guidance.** This option, first suggested by the NSABB (2006), would require that companies obtain a certificate that could be requested by institutions when they make purchasing decisions. The U.S. government would determine the criteria to be used in the certification process.

   Because federal grantees and contractors make up a very large percentage of dsDNA customers in the U.S., such a policy would create a strong incentive for companies to comply with the Guidance and be certified. However, this policy also would favor established companies that already comply with the Guidance and could increase barriers for new entrants into the industry. By providing resources to reduce administrative burden (such as a database of sequences of concern, discussed below), the U.S. government could make this option more acceptable to these less-established companies.

2. **Provide a database of "sequences of concern" for DNA providers to use for screening with guidance on how to evaluate homology.** The NSABB (2006) also recommended that the U.S. government develop standardized "databases and software tools" to aid in sequence screening. The current Guidance is based on whole organisms (e.g., the Select Agent List) and does not prioritize sequences associated with pathogenicity or toxicity. A large percentage of the administrative burden reported by the IGSC is in following up

Ultimately, a more targeted database could reduce the ambiguity associated with bioinformatic screening, reduce the amount of time that companies spend on sequences that are not likely to be used to construct harmful pathogens, and direct resources toward those that could be.

on "yellow" hits, which are sequences that have high homology to pathogenic species but are thought to be harmless or are not related to pathogenicity. (See Box B.) The HHS Guidance itself makes the case that time and effort should not be spent on such "house-keeping" genes, albeit without further definition (HHS, 2010).

A database of this type also could prioritize sequences that, given the state of synthetic biology, could be made pathogenic by moderately skilled lab workers ordering synthetic DNA. Such a database would emphasize pathogenic viruses that are easier to construct (i.e., those with smaller genomes) or are easier to make viable from dsDNA (e.g., positive-stranded RNA viruses, which are easier to make viable than negative-stranded viruses). This database also could be updated with new sequence data and as new threats emerge, including those that are not directly associated with the Select Agent or Commerce Control lists. Ultimately, a more targeted database could reduce the ambiguity associated with bioinformatic screening, reduce the amount of time that companies spend on sequences that are not likely to be used to construct harmful pathogens, and direct resources toward those that could be.[15]

15 This database supports the "Best Match" approach described in the current Guidance and practiced by current dsDNA providers, where providers first determine best matches to a DNA sequence from GenBank and then use this database to check for potential hits. However, a homology (i.e. sequence similarity) approach based on this database alone may be an option in the future as the sequences of concern are better understood and as the number of sequences in GenBank increases to a point that the number of matches becomes untenable.

Expanding the HHS Guidance to Address Other Means of Obtaining dsDNA

One challenge to the effectiveness of the HHS Guidance is the potential for DNA synthesis to become more decentralized and less dependent on the dsDNA synthesis industry. The use of short, single-stranded DNA ("oligos") to build genes and gene fragments is a well-established laboratory technique, and benchtop dsDNA synthesizers are newly available. The increasing use of either of these in the future may undermine the effectiveness of the Guidance.

Oligos and Oligo Synthesizers

The HHS Guidance is directed toward providers of dsDNA and does not address oligos. Over the past five years, the use of oligos to construct genes and gene fragments has become a common laboratory procedure, with commercial kits now available to facilitate this task.[16] Given that gene and viral genome synthesis from oligos is now within the capabilities of far more people than when the Guidance was adopted, the question naturally arises as to whether it now is desirable and feasible to screen orders of oligos.

The lessons learned by DNA providers from screening dsDNA suggest that screening oligos with a similar procedure would be untenable. Because oligos are much cheaper than gene-length dsDNA (often approximately $10 per oligo) and the orders are filled much more quickly (usually overnight), the added costs and time associated with biosecurity screening could be a major burden. (See Box B.) This market is very competitive both within the U.S. and internationally, and profit margins are small. Furthermore, because the sequences are shorter and because oligos are used for a wide variety of purposes beyond gene construction, a bioinformatic screen of a single oligo is likely to yield more ambiguous hits with fewer clues about the researcher's intentions than is seen with dsDNA.

However, it may be possible to separate out orders of oligos that are likely to be used for gene synthesis and focus screening efforts on those. Most oligos are ordered for PCR, quantitative PCR, or sequencing purposes.[17] These oligos generally are short (under 30 nucleotides), and many orders contain only a few oligos. In contrast, oligos that will be used for gene synthesis are generally between 40 and 60 nucleotides in length.[18] Furthermore, many oligos would need to be ordered to construct a gene; a reliable strategy using 60-nucleotide oligos would require 333 oligos to make 10 kilobases of dsDNA (10 kilobases could be a bacterial pathogenic pathway or a

The lessons learned by DNA providers from screening dsDNA suggest that screening oligos with a similar procedure would be untenable. However, it may be possible to separate out orders of oligos that are likely to be used for gene synthesis and focus screening efforts on those.

16 Both NEB (https://www.neb.com/products/e5510-gibson-assembly-cloning-kit) and SGI-DNA (https://sgidna.com/hifi_kit.html) offer Gibson Assembly kits that allow researchers to build dsDNA from multiple oligos in a single reaction. These kits cost from $159 to $185 for ten reactions. Error correction usually is recommended as well; a "CorrectASE" enzyme can be purchased from Life Technologies for $276 for fifty reactions (https://www.lifetechnologies.com/order/catalog/product/A14972).

17 A recent industry report estimated that PCR, qPCR, and sequencing made up 86.9% of the market, in dollars, in 2014 (MarketsandMarkets, 2014) (43.5% for PCR; 31.2% for qPCR; and 12.2% for sequencing). The same report estimated that oligos for gene synthesis represented 5.1% of revenues for the oligo industry.

18 Oligos that are 50-60 nucleotides long are very reliable for gene construction. Oligos above 60 nucleotides are more difficult to synthesize and are often more expensive. Although gene synthesis using 30-nucleotide oligos is possible, the error rate is much higher using these shorter oligos.

small viral genome, such as the 1918 flu). Oligo providers currently offer oligos in 96-well or 384-well plates at a discounted rate.

Taking these factors into account, there are three options that could be pursued to make screening oligo orders more feasible and cost-effective. Note that these options are not mutually exclusive and one or more could be adopted simultaneously:

As new technologies and platforms for dsDNA synthesis are developed, it is quite plausible that benchtop synthesizers will become more widely available. In anticipation of such products, the U.S. government may want to consider ways to incorporate this new capability into the HHS Guidance by adding guidance for providers of dsDNA synthesizers.

- Apply the Guidance only to orders more likely to be used for gene synthesis, for example, those containing many oligos[19] and/or oligos larger than a specified size, such as 30-40 nucleotides. Some other feature of the order (e.g., additional purification to ensure accurate sequence) also could be used as a trigger for screening.
- Develop a much-reduced database of sequences of concern that would include only the sequences that are most pathogenic or toxic and that are easiest to construct from oligos. This database could be some subset of one developed for screening orders of dsDNA (as discussed above) or developed separately.
- Encourage oligo providers to screen their customers so that they are confident that the oligos are going to a legitimate institution or known entity. U.S. oligo manufacturers already have an obligation to screen against U.S. government export control watch lists if they are shipping any of their products to other countries.

Complicating the issue of oligo screening is the increasingly wide availability of both new and refurbished benchtop oligo synthesizers.[20] There is an active international market in oligo synthesizers and they are not identified on any U.S. export control list. It is a challenge to ensure that an oligo synthesizer (especially an older model) will make reliable oligos, and synthesizing oligos more than approximately 60 nucleotides is very difficult. However, with proper maintenance, an oligo synthesizer could be used to generate oligos suitable for gene construction. Given the current lack of government oversight of these machines and their widespread distribution, it is difficult to imagine establishing export controls, registration requirements, or effective guidance for U.S. providers of oligo synthesizers at this point in time.

dsDNA Synthesizers

Until recently, there were no benchtop synthesizers that would make dsDNA. The first such product is SGI-DNA's BioXP 3200, which was launched in April 2015 (SGI-DNA, 2015).[21] As a member of the IGSC, SGI-DNA practices biosecurity screening procedures for orders of dsDNA and has reported that it uses similar practices for the BioXP. Customers will undergo screening, and each order sent through the BioXP will undergo

19 The possibility of a "trusted third party" that could collect oligo orders over time and from different companies has been discussed previously to address the potential concern of "split orders" (Green, 2009). The idea is that this third party (likely the government) would screen the collected set of oligos after they are ordered to determine if any sequences of concern can be constructed. However, such an option is not included here because no agency has expressed any interest in funding such an effort, and it is unlikely that the oligo synthesis industry would voluntarily participate due to confidentiality concerns.

20 A quick Google search for refurbished oligo synthesizers yielded the names of companies in the U.S., Denmark, Russia, and China. As of June, 2015, there were eight oligo synthesizers available on eBay, with prices ranging from $599 to $25,000. Nine more could be found on LabX.

21 Disclaimer: the authors of this report are with the J. Craig Venter Institute, which holds stock in SGI-DNA.

a bioinformatic screen as well. Because SGI-DNA controls the input into the machines (delivered to customers as oligos in barcoded, tamper-proof plates), customers will not be able to produce dsDNA without first sending the sequence to SGI-DNA.

As new technologies and platforms for dsDNA synthesis are developed, it is quite plausible that benchtop synthesizers will become more widely available. In anticipation of such products, the U.S. government may want to consider ways to incorporate this new capability into the HHS Guidance by adding guidance for providers of dsDNA synthesizers. In order to make such an updated Guidance analogous to the current Guidance for providers of dsDNA, manufacturers of dsDNA synthesizers also could be encouraged to screen customers and the dsDNA sequences they will be producing. If federal contractors and grantees are required to purchase dsDNA only from companies that comply with the Guidance (as described above), a similar requirement could be made for purchasing dsDNA synthesizers.

The BioXP represents the current state of the art for benchtop dsDNA synthesis and uses plates of pooled, designed oligos as its input. While it is difficult to predict how future generations of dsDNA synthesizers will work, they will use either oligos or short dsDNA fragments as building blocks. Because current Guidance already covers plates of short dsDNA, additional Guidance is not needed for machines that use them as inputs. Similarly, if oligos used as inputs for dsDNA synthesizers were subject to screening guidance, then separate guidance for manufacturers of those dsDNA synthesizers may not be needed.

References

Carlson R. (2014) Time for New DNA Synthesis and Sequencing Cost Curves. http://bit.ly/sequencing-cost-curves

Carter S.R., Rodemeyer M., Garfinkel M.S., & Friedman R.M. (2015) Synthetic Biology and the U.S. Biotechnology Regulatory System: Challenges and Options. http://bit.ly/syn-bio-report

DeLoache W.C., Russ Z.N., Narcross L., Gonzales A.M., Martin V.J., & Dueber J.E. (2015) An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nature Chemical Biology 11:465. May 18.

Garfinkel M.S., Endy D., Epstein G.L., & Friedman R.M. (2007) Synthetic Genomics: Options for Governance. http://bit.ly/syn-gen-options

Graf M. (2013) "Synthetic Biology: Biosecurity in a Rapidly Emerging Field." Biobricks Foundation SB 6.0, London. July 10. http://bit.ly/syn-bio-biosecurity

Green T. (2009) Hunting Dangerous Genes, Inbox by Inbox. MITRE: Project Stories. February. http://bit.ly/hunting-dangerous-genes

Gibson D.G., Glass J.I., Lartigue C., Noskov V.N., Chuang R-Y., Algire M.A., Benders G.A., Montague M.G., Ma L., Moodie M.M., Merryman C., Vashee S., Krishnakumar R., Assad-Garcia N., Andrews-Pfannkoch C., Denisova E.A., Young L., Qi Z-Q., Segall-Shapiro T.H., Calvey C.H., Parmar P.P., Hutchison III C.A., Smith H.O., & Venter J.C. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science 329:52. July 2.

HHS (Department of Health and Human Services). (2010) Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA. Federal Register 75, 62820. October 13. http://bit.ly/framework-dsdna

IASB (International Association Synthetic Biology). (2009) The IASB Code of Conduct for Best Practices in Gene Synthesis. November 3. http://bit.ly/gene-synthesis-conduct

ICLS (International Council for Life Sciences). (2013) Security Aspects of Synthetic Biology: Meeting Report: 7-8 March, 2013, Hong Kong. http://bit.ly/syn-bio-security-aspects

IGSC (International Gene Synthesis Consortium). (2009) International Gene Synthesis Consortium (IGSC) – Harmonized Screening Protocol – Gene Sequence & Customer Screening to Promote Biosecurity. November 18. http://bit.ly/harmonized-protocol

IGSC. (2015) International Gene Synthesis Consortium Forms Not-for-Profit Corporation. April 28. http://bit.ly/igsc-forms

MarketsandMarkets. (2014) Oligonucleotide Synthesis Market by Product & Services (Equipment, Reagent, Primer, Probe, Custom Oligos), End-User (Research, Pharmaceutical & Biotechnology), Application (Diagnostics, PCR, QPCR, Gene Synthesis, NGS, DNA, RNAi) – Global Forecast to 2019. August. http://bit.ly/oligo-synthesis-market

NIH (National Institutes of Health). (2013) NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules (NIH Guidelines). http://bit.ly/nih-guidelines

NSABB (National Science Advisory Board for Biosecurity). (2006). Addressing Biosecurity Concerns Related to the Synthesis of Select Agents. http://bit.ly/nsabb-biosecurity

SGI-DNA. (2015) SGI-DNA Launches the BioXp 3200 System Early Access Program. April 2. http://bit.ly/sgidna-bioxp

Simirenko L. & Hillson N.J. (2015) "BLiSS: Black List Sequence Screening Pipeline." Synthetic Biology, Engineering, Evolution & Design, Boston. Poster Session B, June 12.

Acknowledgements

We would like to thank the Alfred P. Sloan Foundation for funding this work and Paula Olsiewski, Program Director, for her support.

This project would not have been possible without the insights, perspectives, guidance, participation in our workshop, review of our drafts, and feedback from a wide range of people. We would like to thank the following contributors:

- Board of Directors of the International Gene Synthesis Consortium
- Kavita M. Berger, Gryphon Scientific
- Kellie Bolling, U.S. Department of Energy
- Andrew Bond, Gen9
- Rob Carlson, Biodesic
- Peter Carr, Lincoln Laboratory, MIT
- Isaiah Chuang, Blue Heron/Origene
- Leremy Colf, U.S. Department of Homeland Security
- Susan Collar-Monarez, White House Office of Science and Technology Policy
- Tim Conlon, Integrated DNA Technologies, Inc.
- Jennifer Corbin, Gryphon Scientific
- James Diggans, Twist Biosciences
- John Dileo, MITRE Corp.
- Sarah Edwards, Berberian & Company/U.S. Department of Defense Advanced Research Project Agency (DARPA)
- Gerald L. Epstein, U.S. Department of Homeland Security
- Meg Flanagan, U.S. State Department
- Marcus Graf, Thermo Fisher Scientific
- Gigi Kwik Gronvall, University of Pittsburg Medical Center for Health Security
- Kathryn Harris, National Institutes of Health Office of Biotechnology Activities
- Nathan Hillson, U.S. Department of Energy Joint Genome Institute
- Jeffrey Hung, GenScript
- Kristen Jordan, Intelligence Advanced Research Projects Activities (IARPA)
- Todd Kuiken, Woodrow Wilson International Center for Scholars

- Maio Lu, Genewiz
- Jason Matheny, Intelligence Advanced Research Projects Activities (IARPA)
- Michael Montague, Independent contractor, Intelligence Advanced Research Projects Activities (IARPA)
- Anna Muldoon, U.S. Department of Health and Human Services Office of the Assistance Secretary of Preparedness and Response
- Kimberly Orr, U.S. Department of Commerce
- Chris Park, U.S. State Department
- Ari Patrinos, Synthetic Genomics, Inc.
- Todd Peterson, Synthetic Genomics, Inc.
- André Rusch, Thermo Fisher Scientific
- Reed Shabman, J. Craig Venter Institute
- Dave Shepherd, U.S. Department of Homeland Security
- Howard Simon, International Gene Synthesis Consortium
- Amy Smithson, formerly Monterey Institute for International Studies
- Tim Trevan, International Council of Life Sciences
- Jessica Tucker, National Institute of Biomedical Imaging and Bioengineering
- Damian Urena, Genewiz
- David Walburger, MITRE Corp.
- Edward You, Federal Bureau of Investigation

About the Authors

Sarah R. Carter, Ph.D.

Policy Analyst, J. Craig Venter Institute

Dr. Carter is a policy analyst at the J. Craig Venter Institute where she focuses on societal and policy implications of 21st-century biology. Most recently, she led a project on the U.S. biotechnology regulatory system and the challenges that will arise as synthetic biology and its applications become more prevalent (Carter, et al., 2014). She previously was a policy analyst at the White House Office of Science and Technology Policy (OSTP) where she focused on issues relating to climate change and sustainability. Earlier, she served as a AAAS Science and Technology Policy Fellow at both OSTP and at the Environmental Protection Agency, and as a Mirzayan Fellow at the National Academies. She earned her Ph.D. in Neuroscience in 2007 from the University of California-San Francisco and her bachelor's degree in Biology in 2000 from Duke University.

Robert M. Friedman, Ph.D.

Vice President for Policy and University Relations, J. Craig Venter Institute

Dr. Friedman is Vice President for Policy and University Relations at the J. Craig Venter Institute. He directs JCVI's Policy Center and is also active in several projects ongoing in the Institute's Environmental Genomics Group. Prior to joining the Venter Institute, Dr. Friedman was Vice President for Research at The Heinz Center, a nonprofit policy research organization that brings together collaborators from government, industry, environmental organizations, and academia. Earlier, he was a Senior Associate at the Office of Technology Assessment, U.S. Congress (OTA). For 16 years, he advised Congressional committees on issues involving environmental and natural resources policy. Dr. Friedman received his Ph.D. from the University of Wisconsin-Madison, in Ecological Systems Analysis, concentrating in ecology, environmental engineering, and systems analysis. He is a Fellow of the American Association for the Advancement of Science (AAAS).