GRYPHON
SCIENTIFIC

The Custom Synthetic Nucleic Acid Industry and Biosecurity: A Systems Analysis

DECEMBER, 2007

Prepared For:
The Department of Homeland Security:
Science and Technology Directorate

The Department of Homeland Security:
Office of Policy

Prepared By:
Gryphon Scientific, LLC
Rocco Casagrande, Project Director
Jennifer Byers
Nick Wills
Skye Wills
Yu-Ming Liou
Alicia Isom
Elizabeth Kramer

TABLE OF CONTENTS

Executive Summary

This report is the culmination of a three month end-to-end system study intended to provide information on the trade-offs associated with options for reducing the biosecurity risks associated with the custom, synthetic nucleic acid (CSNA) industry. The conclusions in this report are based upon interviews we performed with CSNA suppliers and customers, along with data from the science policy and molecular biology literature. In our interviews, we were able to obtain information from most CSNA suppliers in the U.S., including all of the largest suppliers, together representing more than 85 percent of the value of the CSNA industry.

A review of the biosecurity policy literature of the CSNA industry reveals an ongoing dispute as to whether or not the CSNA industry should be regulated. While some feel that the availability CSNA poses a great biosecurity risk, necessitating regulation, others, citing the likely benefits of the industry to biomedicine, hold that regulation will inflict more harm than good because the risk of misuse is low. Those calling for regulation are conflicted about who should be in charge and what methods should be used. Most who have written about this subject recognize that there are many issues such as social and economic costs and global implications that must be taken into account when proposing regulation of the industry. Many who support regulation feel CSNA sequences should be screened and/or customers licensed before products are sold. It is clear from the literature available that there is no silver bullet that can prevent misuse of the technology and that any regulation imposed will have a set of costs associated with it. This being said, one must also keep in mind the costs associated with maintaining the status quo. It is with these thoughts in mind that our system analysis was undertaken.

A review of the CSNA industry reveals that it is actually composed of two, easily separable, components: the gene synthesis companies and the oligonucleotide synthesis companies. The gene synthesis industry is relatively small, producing less than 0.2 percent of the volume of products delivered by the oligo industry each year. Many providers of synthetic genes are aware of the potential for product misuse and are therefore open to the idea of screening sequences for pathogenic elements; many are already doing so. Most in the oligo industry, however, do not feel that screening sequences is necessary or practical due to the small size of their product and the high-throughput nature of the industry. Despite this opposition of the oligo industry, it appears there is a 15-60 minute window between the time orders are placed and synthesis begins where sequence screening could be introduced. Both synthetic gene and oligo manufacturers are amenable to the idea of screening customers to ensure products are only shipped to those with legitimate ties to the scientific community and many in both industries are already screening customers. Providers of both synthetic genes and oligos have expressed their desire for a designated contact within the government where suspicious customers or orders can be reported. With the CSNA industry already valued at over $1 billion and growing, representatives from the industry hope to prevent the misuse of technology without substantial costs to their business or customers.

Several existing regulations partially address the risk associated with CSNA. Although the Select Agent Rules (SAR) provide a robust framework for the control of the pathogens themselves, at present, they are generally understood not to apply to fragments of a genome of a pathogen or fragments of a gene encoding a toxin, or potentially even to an intact genome of viruses that need additional components to enable expression. Furthermore, the quantity of researchers regulated by the SAR (approximately 8,000 registered with the CDC and 6,000 with the USDA) is far fewer than those ordering CSNA (~200,000), complicating the direct application of procedures from the SAR to this field. EPA regulations under the Toxic Substances Control Act (TSCA) do not apply to naked nucleic acid, but would apply when the nucleic acid is inserted into an organism. Where regulations under TSCA are relevant to CSNA are the mechanisms used to control and protect confidential business information on biologicals. Commerce Control Regulations *should* apply to a significant fraction of the several million CSNA orders exported

yearly, but, in practice, CSNA companies do not apply for export licenses for several reasons: the unit within the Department of Commerce responsible for the oversight of dual-use exports lacks the capacity to vet these orders, CSNA companies are largely ignorant of these regulations, the time required to obtain a license is far greater than the usual delivery time of CSNA products, CSNA suppliers may not know their product is dual-use and, due to the difficulty of properly identifying CSNA products in a package, it is impossible to enforce. Because the quantities of researchers using radionuclides and CSNA are roughly similar, mechanisms to license and regulate the use of radionuclides may be instructive to the oversight of CSNA. Robust oversight of radionuclides clearly requires on-site visits with potential licensees. The EU is currently considering licensing life-scientists to prevent hostile actors gaining access to useful reagents and knowhow.

Several critical choices face those considering an oversight regime for the CSNA industry. The first choice is fundamental: which products of the CSNA industry should be regulated? Decisions here identify the critical balance between burdens to industry, burdens to researchers, and potential risk reduction. In order to prevent hostile actors from synthesizing genes and viral genomes de novo, the system should regulate genes and oligos greater than 35nt in length. In order to prevent the amplification of genes from medical or environmental samples via PCR by hostile actors, the system should include oligos of any length. If only long oligos (>35nt) and genes are included in the system, most of the CSNA sold daily, most customers and some CSNA producers would not need to be included in the system. Also, because short oligos are more numerous and more likely by chance to match a suspect sequence, the exclusion of short oligos from the system is likely to dramatically reduce the false hit rate in primary screening. Furthermore, CSNA products can be DNA analogs or chemically modified DNA, or come attached to substrates. Most of these non-standard-DNA products are not ideally suited for use in PCR or for viral or gene synthesis and would only be used in the illicit production of a biological agent if the weaponeer feared identification by the oversight system. Excluding non-standard oligos would have a minimal benefit on reducing the total quantity of orders that must be screened, but would have the benefit of putting several small companies outside of the oversight regime. Final decisions on which CSNA types and formats should be beyond regulatory reach should be based upon a comprehensive weapons-process assessment.

The second main question to resolve is: what group will provide oversight and perform the screening? Screening performed by the CSNA providers themselves would not only be more palatable to the industry itself due to the retention of proprietary information about CSNA orders but also because the system would likely screen efficiently enough, due to the fact that the screeners and suppliers are co-located, to enable the supply of oligos in a time-frame expected by their customers. In contrast, screening performed by a single, national entity would ensure compliance, be less likely to be circumvented by an adversary and facilitates the use of data from ongoing screening to improve future screening. However, this system may be too cumbersome to meet the time demands of the oligo industry and would require the transmittal of proprietary business information.

There are several options for screening an order, including screening the customers only or screening customers in the context of the sequences they order. Because we estimate that approximately 1,000 orders are made daily by legitimate researchers for CSNA sequences associated with pathogens (including both controlled and uncontrolled pathogens), customer information must be included in any screening system. Customers could be screened by information in their e-mail address, by screening against lists of known terrorists or denied parties or by ensuring that the customer possesses a license to obtain CSNA. These measures would be effective at preventing hostile actors from acquiring critical reagents but would have little biosafety benefit. In contrast, if licensing is combined with sequence screening, some customers could be issued licenses to obtain CSNA from pathogens; those without this type of license could only obtain CSNA from non-pathogenic organisms, thereby improving biosafety. A licensing system could be implemented in a variety of ways, including grandfathering in those who have

ordered CSNA products in the past. However, it is unclear how licensing would be implemented for foreign customers of US companies. Sequence screening could be performed by software that is currently available, such as BlackWatch. This software could be improved or entirely new software could be developed. Details on what CSNA orders should be screened against must rely on a detailed weapons process assessment. Any hit should be confirmed before an order is denied or a field investigation is initiated. This confirmatory screening could involve a phone call to the customer's biosafety officer, or determining if the sequences ordered match their previous orders better than pathogens of concern.

Although it is clear from existing practices that questions of biosafety should be investigated by the CDC or USDA, questions of biosecurity could be investigated by the FBI and/or DHS's Homeland Infrastructure Threat and Risk Assessment Center (HITRAC). No matter who performs the investigation, standard practices need to be immediately implemented even if no single screening system is adopted. CSNA producers must be supplied with a point of contact at the USDA-Animal and Plant Health Inspection Service, CDC, the FBI (the WMD coordinator of the local field office) and/or HITRAC. Furthermore, standard procedures must be given to the CSNA supplier as to whether or not to fill the orders referred to a field investigation, and if the order is not to be filled, what to tell the customer.

Furthermore, to support forensic investigations, information on orders should be stored. If these orders are stored at the CSNA producers and only inspected when needed, forensic investigations would be slowed but proprietary information would be protected. If order information (or information only on denied orders or orders that failed primary screening) is stored in one location, this database would be easier to search in times of need but may raise concerns dealing with proprietary business information.

Finally, a CSNA oversight system could be underpinned by a variety of mechanisms. Self-regulation is, by its nature, the most palatable to industry, but is likely to be inconsistent and leave gaps to be exploited by adversaries. Voluntary federal standards could be adopted to improve adherence to the system industry-wide, and may be palatable to industry because of their desire for guidance from the federal government and the possible reduction in vulnerability to tort claims. Guidelines that dictate that any institution receiving NIH (or other government) funding can order CSNA only from suppliers that employ a standard screening system would improve compliance to the system by industry in the US and abroad by controlling the buying patterns of a large part of the customer base. Furthermore, companies could not escape regulation by relocating production facilities overseas. Lastly, screening could be compelled by formal changes in US law or regulations, such as the Select Agent Rule and the Commerce Department Regulations.

It is important to consider that the most effective oversight system may actually be multiple systems in one: one tuned to provide oversight to the gene industry and one to the oligo industry. We realize that the requirements of the customers of the gene synthesis, compared to the oligo synthesis, industry are vastly different as are the time and financial constraints faced by both industries. Furthermore, an oversight system may be implemented by including components of the CSNA industry one at a time in a phased implementation approach. Because many gene synthesizers already perform screening, because the quantity of genes made pales in comparison to the quantity of oligos made, because genes are probably more enabling to adversaries than oligos, and because the cost and time constraints of gene production are more relaxed than oligo production, implementing regulation by starting with the gene synthesis industry seems appropriate.

The cost of a CSNA screening system could almost certainly be covered by fees attached to CSNA products. From talking to customers, we estimate that the market would be able to raise approximately $100 million a year in user fees without disrupting the buying patterns of customers. This amount of funding would support most types of screening systems except ones that exactly mirror the rigors of the Select Agent Rule.

Acknowledgements

This report was researched and drafted in the three month period from May 15 to August 15, 2007 and was funded by the Science and Technology Directorate and the Office of Special Projects (Identification and Assessments Division) of the Department of Homeland Security.

We appreciate the efforts of Dr. Sara Klucking, the Contracting Officer's Technical Representative on this project, for her guidance and advice during this project. Her insight helped make this project successful despite its extremely ambitious timetable.

The authors would like to thank our consultants and advisors who either critically reviewed the document or provided advice and guidance. We thank Dr. Ken Oye and Joan Mullen for critical review of much of the document, Dr. Molly Pickett for providing data related to industry requirements and limitations and Joe Chuchla for providing advice on the role of the Commerce Department on the regulation of exportation of dual use items.

Chapter 1: Introduction

Purpose

The advance of biotechnology has not only ushered in remarkable discoveries that benefit mankind, but also has shifted some technological barriers to the acquisition of potentially dangerous pathogens. It is feared that our adversaries could secretly employ companies that send custom-ordered genes and/or pieces of DNA, oligonucleotides—or oligos, through the mail and use these reagents to circumvent existing regulations that prevent the acquisition of pathogens by the unlicensed. To this end, the Department of Homeland Security, Office of Special Projects—Identification and Assessments Division, the Office of Policy, and the Chemical and Biological Division has sponsored Gryphon Scientific to study options for a system that would screen orders for synthetic oligos and genes.

The purpose of this study is not to present a single, best system that we suggest should be adopted by the industry and the government, but to present options for regulation (or the absence of regulation) and the laws and concepts of operation (ConOps) that support it. As background, we also present an analysis of relevant existing regulations. Given the ambitious schedule of this effort, we necessarily cannot explore all aspects of all system options fully.

The NSABB is currently drafting recommendations for controlling dual-use biological technology, and issued recommendations on synthetic genomics in October 2006. In this report, we further expand upon the recommendations already presented by the NSABB, and include information about the costs, benefits and feasibility of many different options for limiting the risk posed by the unregulated access to custom nucleic acids.

We, and our sponsors, recognize that the efficient operation of the industry is essential to the progress of industrial and academic science in the US and the world. Furthermore, the United States is the world leader in synthetic oligo and gene production, and we recognize the possibility that burdensome regulation could push these companies overseas. For these reasons, the impact of regulation on the industry was carefully considered and presented.

A Note on Terminology

In this report, we refer to the genes and oligos made to order by industry by the somewhat awkward acronym CSNA (Custom Synthetic Nucleic Acids). This term avoids inaccuracies that would crop up if we referred to all synthetic genes as oligos or vice versa. Therefore, in the text when the word “gene” is used, we are discussing only genes (long, double-stranded pieces of DNA made in a multi-step process that uses oligos) and when the word “oligo” is used, we are referring only to oligos (shorter, single-stranded pieces of DNA that may or may not have been pieced together from smaller oligos). Similarly, in contrast to the term “synthetic DNA”, this term accounts for the fact that a substantial component of the CSNA industry is comprised of custom RNA oligos, although we do mean to include in this term other nucleic acid analogs that are not technically nucleic acids, like morpholinos.

We use various terms to describe the length of a CSNA product. Often, the term nucleotides (abbreviated “nt”) is used to describe the length of a single-stranded oligo. Base pair (abbreviated “bp”) is used when describing the length of double-stranded DNA, where each nucleotide is paired with its complementary base. We use the term “base” and “nucleotide” interchangeably in this proposal (even though this is biochemically inaccurate as base refers to a component of the nucleotide) to describe the subunits of which an oligo or gene are composed. When talking about long stretches of DNA, we use standard

scientific terminology to describe increasing orders of magnitude (kbp for 1,000 base pairs, and Mbp for 1,000,000 base pairs).

Furthermore, we often avoid the use of the terms "false positive" and "false negative" in this report due to the fact that these terms are most appropriately used in the context of a system-wide event. In this report, we use the term "false hit" or "false miss" to describe when a system component identifies a problem that poses no real risk, or misses a problem that truly poses a risk, respectively.

Scope

Although it could be argued that many biotechnologies are changing the balance between biological offense and defense, none has had its potential demonstrated as publicly as the ability to rapidly and cheaply obtain CSNA. The synthesis of poliovirus (and the later, more rapid synthesis of a bacteriophage) demonstrated that oligos can be used to obtain some viruses without access to environmental or medical samples that contain the viruses. Furthermore, it has long been appreciated that molecular techniques, such as the polymerase chain reaction (PCR), can be used to amplify and acquire genes encoding toxins if an appropriate sample of pathogen-containing material is available. In Chapter 2, we discuss in greater detail both the promise of the increasing capability of DNA synthesis, and its potential pitfalls. Because of its prominence amongst other aspects of biotechnology, this report will focus exclusively on the CSNA industry. This report will not examine other aspects of biotechnology with potential impact on biosecurity.

For this reason, we will not discuss the regulation of the synthetic biology field, in general. To be clear, the field of synthetic biology involves the manipulation of life using synthesized biological macromolecules. In other words, synthetic biology can be used to create forms of life that have not existed before in any combination. In contrast, the activities we are considering in this report are simply the recreation of life that already exists through chemical synthesis, or the borrowing of a particular phenotype of an existing organism and imbuing it into a different organism insofar as these activities are enabled by the purchase of CSNA. In this report, we are not considering that a handful of rogue scientists will outpace the best minds working feverishly at the most prestigious institutions across the country by applying synthetic techniques in new ways to create new life. We simply are concerned with hostile actors copying what has already been done in the published literature with different species and with malicious intent.

Given the breadth of what synthetic biology encompasses, this field is imbued with greater promise and complications than the synthesis of nucleic acids. Of course, CSNA are essential reagents needed for most efforts in synthetic biology, and therefore, the costs and the benefit of oversight of CSNA and oversight of synthetic biology, writ large, cannot easily be disentangled.

Although more attention recently has been paid in the literature to the de novo synthesis of viruses, we have attempted to discuss all activities of biosecurity concern that could be attempted with CSNA. Similar techniques could be used to directly synthesize genes that encode toxins or other pathogenic components to imbue less innocuous organisms with pathogenicity (or for the harvesting of toxins). Thinking further into the future, an entire bacterial genome could be synthesized and transferred into a closely related, but easily obtainable, bacterium to indirectly obtain a pathogen of interest. (In fact, the Venter group recently transmuted one bacterial species into a closely related species by introducing the genome from the closely related species, in this case harvested from the closely related bacteria, into the original bacteria.[1]) Furthermore, in addition to considering the synthesis of viruses (and similar techniques), we explicitly discuss system options for the oversight of products useful for the PCR-based amplification of genes for

[1] Lartigue, C. et al, "Genome transplantation in bacteria: changing one species to another" *Science*, **317**(5838), 2007.

toxins and other pathogenic components from environmental or medical samples. We present data that will enable a trade-off analysis to determine if the extra costs associated with a system that includes the screening of reagents need for PCR (as compared to a system that just screens reagents needed for gene and virus synthesis) are worth the benefit derived.

When considering biosecurity issues related to CSNA, it is important to note that CSNA could be obtained either by placing an order with a supplier or by synthesizing the desired CSNA oneself. In this report, we consider only the supply of CSNA by industry, not those made with a DNA synthesizer owned by the group wishing to acquire the CSNA. In this way, this report focuses on the simplest, most foolproof and most rapid method a group could use to acquire needed oligos. Furthermore, the challenges associated with oversight of the industry that makes CSNA (vs. the oversight of the use of synthesizers that have been on the market for years or are themselves buildable) is tractable. That being said, we envision that the options for oversight of the oligo industry laid out in this report will be useful as a framework around which further rules for the oversight of the synthesis of oligos in general could be built.

If a nationwide screening system prevents a weaponeer from obtaining needed CSNA reagents from US companies, he could still obtain CSNA products from foreign companies. For this reason, for a screening system to be effective, it must encompass CSNA providers internationally. We expect that a future phase of this project will focus on creating an international regime based on the screening system deemed acceptable by US government and industry. In this report, we have focused on efforts to get an effective screening system that would prevent a weaponeer from obtaining CSNA products from US companies. We have only interacted with foreign CSNA providers in order to collect best practices related to screening.

Also, because it was not the purpose of this report to provide details on risk reduction metrics, it was not necessary to perform a complete technical weapons-process assessment related to exactly what CSNA would be enabling in the hands of an adversary (and what associated technologies were needed). However, in this report, we present a brief technical assessment to gauge some preliminary costs and benefits of including or excluding various oligo products (modified oligos, DNA analogs and lengths of oligos) in the oversight systems and the tradeoffs associated with the screening of CSNA of various lengths. A thorough weapons-process assessment will be required to identify how exactly a sequence screening system should work, the exact products that should be included in an oversight system, and the pathogen sequences that we do not wish an adversary to acquire (many viruses may be too difficult to synthesize from CSNA as they may need accessory proteins introduced along with the genome).

Furthermore, because this project had to be completed in three months, we decided to touch on as many issues related to this topic as possible instead of presenting extensive detail on only a few issues. Since the purpose of this report is to present the costs and benefits of options for oversight to policymakers, we assume that this approach is superior to providing details on costs or implementation for only a few options.

Approach and Organization

This report is divided into five chapters, the first of which is this introduction. Chapters 2, 3, and 4, the Literature Review, the Industry Review, and the Regulatory Review, respectively, provide data needed to understand the costs and benefits described in Chapter 5, which details the trade-offs in the notional system.

Chapter 2 is the Literature Review, in which we abstract the community's thought on the biosecurity and the CSNA industry. It is important to note that many commentators who have made remarks relevant to the regulation of the CSNA industry were, in fact, focused on an industry encompassing more than the CSNA industry, the synthetic biology industry. In synthetic biology, researchers attempt to fabricate organisms that heretofore have not existed or synthesize existing microbes completely de novo. Our report, and the CSNA industry, concerns itself simply with the synthesis and delivery of custom-made pieces of nucleic acid. Clearly, the synthetic biology industry is heavily dependent on the CSNA industry for critical reagents and the risks associated with CSNA stem from the fabrication of existing microbes de novo or the manipulation of existing life to create a strain that has heretofore not existed. Therefore, in the literature review we consider the literature on synthetic biology as it applies to the CSNA industry.

Chapter 3 is the Industry Review, in which we provide data on the CSNA industry. This chapter itself has several subsections, each of which is drawn from different data sets. The first subsection describes the products of the CSNA industry and is drawn from a large commercial survey of CSNA users, our interviews with CSNA suppliers and our experience with molecular biology. This section describes how types of CSNA products differ from each other and how the risk of misuse of each product interplays with the need to include each product in an oversight system. The second subsection discusses the CSNA suppliers themselves, and is drawn almost exclusively from interviews with all of the major CSNA suppliers in the US. The third subsection discusses data about the customers of the CSNA industry, and, importantly, their beliefs and opinions regarding oversight of the industry. The data in this section was drawn from a large commercial survey of CSNA customers and from further detailed interviews we conducted with life-science researchers. The last subsection of the Industry Review discusses the approaches and tools used to investigate CSNA sequences for suspicious content. The data in this section were obtained from the scientific literature, interviews with developers and statistical analysis.

Chapter 4 describes the regulations, laws and guidelines that are relevant to the oversight of the CSNA industry. The data in this section are drawn from the literature, interviews with experts in the field and the text of the regulations and guidelines themselves. Importantly, we also discuss regulations that control products similar to CSNA as well as those that are directly relevant.

Chapter 5 of this report considers all of the data discussed in the first four chapters and describes the trade-offs for each component of a notional system that provides oversight of the CSNA industry. Because this section is based upon the data presented in the first four chapters, we attempt to recapitulate data already presented as little as possible.

Chapter 2: Literature Review

While the first complete *de novo* synthesis of a gene was achieved more than 35 years ago, by Khorana and colleagues[2], the techniques and tools available for synthesis remained primitive until recently, making synthesis a difficult and arduous task that often took years to complete. However, with the advent of commercially-available DNA synthesis machines, individual bases can be assembled quickly and easily, in any specified sequence, using inexpensive and accessible reagents. While individual laboratories can readily purchase both new and used synthesis machines, researchers more commonly order relatively short DNA segments from commercial oligonucleotide synthesizers. Longer, gene or even genome-length sequences can be purchased from gene synthesis companies. The easy availability of DNA synthesis offers, as the authors of the Sloan Foundation Report note, "the potential for revolutionary advances, making possible qualitatively new capabilities", and crucially, "broadening the number of scientists and engineers able to use biotechnology, and enabling them to consider higher-level applications".[3] However, the advent, development, and spread of CSNA technologies have also led to concerns that bioterrorists could take advantage of these techniques to construct bioweapons from scratch. Garfinkel et al. argue in the Sloan Foundation report that "synthetic genomics…is [thus] a quintessential 'dual-use' technology—a technology with broad and varied beneficial applications, but one that could also be turned to nefarious, destructive use."

In the following section, we review attempts within the existing literature to address the questions being asked about the control of the synthetic oligo industry: Should there be oversight of the synthetic DNA industry? Who should be in control? How should the oversight regime operate and what are the needs, requirements, and limitations that should be considered when proposing regulation?

Should there be Oversight of the Synthetic DNA Industry?

Promise of the CSNA Industry

When assessing the costs of regulating the synthetic DNA industry Vriend reminds us that "the development of synthetic biology technologies must be evaluated for their broader socioeconomic, cultural, health and environmental implications and not simply for their misuse in the hands of 'evildoers.'" With that in mind it is important to remember the benefits associated with the technology. According to the DNA for Peace report, "The world must not let legitimate concerns about biosecurity undermine the promotion and use of biotechnologies for human development."[4] In considering governance options for the synthetic biology community, it is vital that policymakers strike a balance between minimizing safety and security risks presented by CSNA and permitting and, indeed, encouraging its ongoing and future development. As Garfinkel et al. stress in the Sloan Foundation-funded Report "Synthetic Genomics: Options for Governance", synthetic genomics not only presents potentially dire risks, but equally, if not greater promises of benefits and advances.[5] In particular, the authors identify several major areas in which the new field of synthetic genomics, that relies upon the CSNA industry, has the potential to make "unique or significant contributions" to scientific discovery, medical treatments, and economic growth.

---

[2] Khorana, HG, et al. "Total Synthesis of a Gene" *Science,* **203** (614), 1979.
[3] Garfinkel, MS, et al. "Synthetic Genomics: Options for Governance". Nov. 27, 2006.
[4] Candian Program on Genomics and Global Health, "DNA for Peace: Reconciling Biodevelopment and Biosecurity". November, 2005.
[5] Garfinkel et al, 2006.

Benner and Sismour point out a few of the many accomplishments in which the CSNA industry has had a hand.[6] Although many of the advances cited in the report cannot be attributed solely to the ability to obtain custom-designed sequences of DNA specifically (as opposed to the ability to manipulate life more generally, which is often called by the broader term "synthetic biology"), some of the examples illustrate the promise of this specific industry. One such accomplishment is the introduction of synthetic diagnostic tools such as Bayer's branched DNA assay which annually helps improve the care of some 400,00 patients infected with HIV and hepatitis. The Sloan report also made note of the many beneficial contributions synthetic biology has made to society.

The authors argue that synthetic genomics can, and already is, changing the nature of basic molecular biological research by several means.[7] "As an enabling technology, it has already been shown to be a significant time saver by shortening the time needed for normally arduous recombinant DNA techniques; in the coming five to 10 years" it should begin to offer significant cost savings as well. Moreover, the use of CSNA to rapidly transform the sequences of various genes or entire genomes is swiftly becoming a powerful driving force behind advancements in a wide range of disciplines. "For example, various laboratories are already using [it] to understand the mechanisms of evolution at the molecular level, to rapidly define regulators of specific genes or gene pathways, and to begin to validate, at the [nucleic acid] level, the minimal requirements for life."

Second, the heightened ability provided by CSNA to manipulate genetic sequences may also enable more efficient research, development, and production of human and animal vaccines, biodefense detectors, or diagnostic tools. In particular, as noted in the Sloan Foundation Report, "the ability to assemble and mutate sequences rapidly could allow for the development of broadly protective vaccines against, and diagnostics for, viruses that [are themselves] diverse and variable, such as SARS and hepatitis C."[8]

The vast potential benefits offered by synthetic genomics must consequently be kept in mind by any would-be regulators lest they unintentionally stifle potentially beneficial research efforts. Indeed, it is this fear that has motivated some authors to argue against government regulation of the CSNA community, suggesting that "it is questionable…whether such efforts will increase security or benefit the public good" and urging policymakers to "resist the impulse to restrict research and the flow of information."[9] Moreover, as Gesteland et al. note in "Synthetic Genomes: Technologies and Impact", while new synthetic genomics technology may serve to enable the synthesis of pathogens, thereby compromising biosecurity efforts, it may just as easily "provide…means for developing vaccines against pathogens, methods for pathogen detection and new ways to identify therapeutic targets,"[10] potentially enhancing current biodefense systems. Gesteland's statement is tempered by its assumption of benefits that may not be realized as compared to a risk that is present today.

Perils of the CSNA Industry

Some objections to additional regulation of synthetic biology have been raised on the grounds that synthetic biology does not represent a radical transformation of biology, but rather an evolutionary development of "classical" recombinant genetics, in which existing genetic elements are harvested from nature and manipulated. These commentators point out that "much of what is currently called synthetic biology is congruent with the recombinant DNA technology" first developed over 30 years ago and argue

---

6 Benner, SA. and Sismour, M. "Synthetic Biology", *Nature Genetics*. July 2005. (6) 7.
7 Garfinkel et al, 2006.
8 Garfinkel et al, 2006.
9 Carlson, Robert. "The Pace and Proliferation of Biological Technologies". *Biosecurity and Bioterrorism: Biodefense Strategy, Practice and Science*. September 2003. (1) 3.
10 Gesteland, R, et al, "Synthetic Genomes: Technologies and Impact", *DOE BERAC*, December 2004. Available at: http://www.science.doe.gov/ober/berac/SynBio.pdf.

that "placing a new name on an old technology does not create a new hazard."[11] Furthermore, they note that regulation of the direct synthesis of genomes will do little or nothing to prevent hostile actors from obtaining pathogens from biological repositories such as the American Type Culture Collection (ATCC)[12] or from environmental samples[13] as in the 1984 salmonella bioterror attacks in Oregon.[14] However, as the ability to rapidly purchase long oligos is a relatively new capability with regulatory implications that are as yet unclear, it may have the effect of providing the terrorist, criminal, or simply irresponsible biologist with a way to make an "end-run" around the existing regulatory framework (such as the Select Agent Rules) covering recombinant genetics and the acquisition of cultures from biological repositories.

In the first case, as Bügl et al. note in a recent (June 2007) commentary in Nature Biotechnology, "synthesis allows the physical decoupling of the design of engineered genetic material from the actual construction and resulting use of the material; DNA can be readily designed in one location, constructed in a second location and delivered to a third."[15] More fundamentally, synthetic genomics may simply provide an alternate route for hostile actors, perhaps lacking the clearance to receive select agents through the normal channels or the expertise or opportunity to collect pathogens from environmental samples, wishing to obtain pathogens for the purpose of causing harm, as Bügl et al., Benner and Sismour, and others[16] have pointed out.

Moreover, although genetic manipulation is certainly possible through traditional techniques, the direct synthesis of custom-tailored genes makes it more accessible.[17] As Gesteland et al note in "Synthetic Genomes: Technologies and Impact", "while the recent synthesis of a viral genome does not provide fundamentally new technology, the new methods used could lower some cost or time barriers" in the development of potentially more potent biological agents. Former Soviet bioweapons scientist Serguei Popov echoed these conclusions in a 2006 interview with *Technology Review*, in which he explained that while in the past, he "had fifty people doing DNA synthesis manually, step by step" to produce weapons-grade viruses, modern DNA synthesizers would dramatically lower the time and cost of producing equivalent results. Whereas 25 or 30 years ago, he explains, "one step was about three hours…today, with the synthesizer, it could be a few minutes—it could be less than a minute. Nevertheless, already the idea was that we would produce one virus a month."[18] While current terrorists or criminal groups are unlikely to possess the scientific resources, whether in the form of technical expertise or facilities and equipment, of a major state-funded bioweapons program like that of the former Soviet Union, a casual extrapolation from Popov's estimates produces chilling conclusions. In the eyes of Kwik et al., the incredible potential for harm presented by such a reconstituted pathogen to a modern, globalized, and open society "warrants vigilance even if the likelihood of bioterrorism is low."[19]

Michele Garfinkel and the other authors of the Sloan Foundation Report argue that potential bioterrorists are most likely to turn to synthetic genomics as a means to obtain certain categories of pathogenic viruses

---

[11] Benner, SA. and Sismour, M. "Synthetic Biology", *Nature Genetics*. July 2005. (6) 7.

[12] The validity of this particular charge is uncertain as the transfer of dangerous (in this case listed) pathogens by entities such as ATCC are currently regulated under the Select Agent Rules, which are considerably more restrictive than any future regulatory system for synthetic biology is likely to be.

[13] Garfinkel et al., 2006.

[14] Kwik, G, et al. "Biosecurity: Responsible Stewardship of Bioscience in an Age of Catastrophic Terrorism", *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*. January 2003. (1) 1.

[15] Bügl, H, et al. "DNA Synthesis and Biological Security", *Nature Biotechnology*, June 2007. (25) 6.

[16] Maurer, S M, et al. "From Understanding to Action: Community-Based Options for Improving Safety and Security in Synthetic Biology", April 2006.

[17] Maurer, SM, et al, 2006.

[18] ETC Group. "Extreme Genetic Engineering: An Introduction to Synthetic Biology," January 2007. Available at: http://www.etcgroup.org/en/materials/publications.html?pub_id=602

[19] Kwik, G, et al. 2003.

(viruses, being simpler than bacteria and possessing much smaller genomes that are more easily synthesized) that are not readily accessible by other means. They concluded, "after examining the viruses on several lists of microbial agents" deemed to "pose severe threats to public health and safety" that the "best" candidates for de novo synthesis from the perspective of a bioterrorist would include "smallpox, filoviruses such as Ebola and Marburg, 1918 influenza virus [also known as the 'Spanish flu'], and perhaps the 2003 SARS coronavirus."[20] Smallpox, 1918 influenza, and the 2003 SARS coronavirus no longer exist in nature and would consequently be impossible to isolate from environmental samples, while the natural reservoirs of Ebola and Marburg were until recently unknown and are not readily accessible. Access to all four of these virus types is, moreover, strictly controlled. Smallpox is known to be held only in two highly-secure laboratories, one in the United States (at the CDC in Atlanta, Georgia) and one in Russia (at the State Research Center of Virology and Biotechnology in Novosibirsk),[21] the Ebola, Marburg, and SARS viruses, are held in only a few tightly-controlled laboratories due to their virulence, and 1918 influenza, which until recently did not exist at all, was "resurrected" by synthesis from DNA obtained from a variety of historical samples. However, the sequences of each pathogen have been determined and are available in public databases.

Garfinkel et al posit that an aspiring bioterrorist seeking to synthesize one of the four previously described types of viruses, would be presented with three options. "The simplest would be to order large stretches of their genome (thousands to tens of thousands of bases long) directly from a gene synthesis company. Much of the hard work needed to synthesize an infectious virus would be done by the company, greatly simplifying" the task. However, this route is the most likely to elicit suspicion and is, as the authors note, the most likely target of government regulators. Should this option be unavailable or deemed too risky, the malefactor might then attempt to order the sequences piecemeal, from a commercial DNA oligonucleotide supplier and assemble these shorter pieces himself. Demonstrating this possibility, a research group at SUNY Stony Brook led by Eckhard Wimmer, a leading polio virologist, caused a substantial stir in August 2002, when a paper announcing their synthesis of a functional version of poliovirus from commercially-ordered DNA oligonucleotides was published in *Science*. Wimmer and his team were criticized in some quarters for potentially inspiring and enabling would-be bioterrorists. In a July 2006 interview with the *Washington Post*, Wimmer told the paper that the experiments had in fact, been undertaken as "a wakeup call" to the synthetic genomics community and revealed that his group had repeated the poliovirus reconstruction a further six times since 2002 and found the process easier and faster with each iteration.[22] While poliovirus has an extremely small genome (7,741 base pairs) and would not be considered a particularly effective bioweapon, Ebola, as Jonathan Tucker notes, is not much more complicated, with a genome that is about 2.5 times larger (though, it should be noted that the synthesis of Ebola would be further complicated by the fact that its genome is coded in negative sense RNA, rather than DNA or positive sense RNA, like the polio virus).[23] Although Tucker concedes that, "given the major technical hurdles involved in the test-tube synthesis of large viruses such as smallpox…the availability of custom-made DNA is unlikely to pose a major threat for the foreseeable future." This apparent contradiction probably indicates that Tucker considers viruses with smaller genomes to be more amenable to CSNA-enabled techniques in the near future than viruses with larger genomes.

Others however, feel the threat is more immediate and that advances in technology may put the synthesis of even more complex viruses like smallpox (with a genome of nearly 200,000 base pairs) within reach. Not only will scientists, whatever their intent, be able to synthesize ever-more-complex microbes, but

---

[20] Garfinkel et al., 2006.

[21] Enserink, M. "Unnoticed Amendment Bans Synthesis of Smallpox Virus", March 11, 2005. (307) 5715; although many scholars also discuss the possibility that smallpox is held in a few, clandestine, repositories by other states.

[22] ETC Group, 2007.

[23] Tucker, JB. "Biosecurity: Limiting Terrorist Access to deadly pathogens", *Peaceworks* United States Institute of Peace, November 2003. No. 52.

they will be able to do it with increasing speed and ease. Indeed, within little more than a year of the Wimmer group's publication came the announcement from the Venter Institute of the synthesis, from commercially available oligonucleotide sequences, of the (5,386bp) bacteriophage (a virus that infects bacteria) phi X 174.[24] What was particularly stunning about the accomplishment was that the Venter Institute team had accomplished it in just 14 days of frenetic work. It should, however, be noted that the planning of the experiment and the design of CSNA reagents was done prior to this period. While the team enjoyed access to extensive resources and included leading scientists such as Nobel Laureate Hamilton Smith, the techniques and technology they used are diffusing steadily throughout the scientific community.[25] It is certainly not fantastical to imagine these techniques being used by future bioweaponeers. "The shoulders of giants" have, after all, borne all manner of men.

Already, as Garfinkel et al. note, databases of DNA sequences and the computation-based design tools used in CSNA synthesis are "rapidly growing in size and sophistication, and are available to virtually anyone in the world with a computer."[26] Thus, they conclude, given access to the appropriate genomic sequence, the only additional necessary ingredients for an individual to re-construct a deadly pathogen "are access to a moderately well-equipped biology laboratory, the know-how to carry out the necessary synthetic techniques, and the biological expertise to take strands of nucleic acid and 'boot' them into viable, self-replicating organisms (or transfer them into a cell so that they express their particular gene products)."[27] Although the level of scientific expertise required is substantial, requiring much greater expertise than the amplification of traditional agents, the use of commercially available DNA sequences would further ease the process, allowing any adversary to simply stitch together pre-fabricated gene segments and to forego the additional effort and expense of operating his own synthesizer.

Although there is some disagreement over the magnitude of the threat posed by the misuse of the technology, it is difficult to dispute the fact that there is a real risk in unregulated access to CSNA. It is with such risks in mind that commentators overwhelmingly conclude that some sort of regulatory scheme for CSNA has become necessary. While there is substantial disagreement about the shape that such regulation should take, it is difficult to find authors willing to reject any attempt to regulate the synthetic biology community. As the drafters of the Sloan Foundation Report argue, to do nothing as the technology of synthetic genomics continues to advance, is not only to do nothing in the face of emerging threats, but to effectively eviscerate existing regulations through inaction as well. If the Select Agent Rules are intended to "prevent unauthorized users from acquiring [listed] pathogens through diversion or theft from legitimate facilities", they are rendered irrelevant by synthetic genomics "to the extent that Select Agents [could] now be synthesized directly."[28] Policymakers may consequently find additional oversight necessary simply to prevent the erosion of the protective capacity of the Select Agent Rules.

We further note that the timeframe considered by commentators in field differs when discussing the perils and benefits of the CSNA industry. Normally, the benefits identified by reviewers will be realized at an indefinite point in the future, whereas the experiments of concern may be occurring now (and have already occurred in the labs of well-meaning scientists). Given this additional consideration, it is clear why that almost all authors conclude that some form of oversight of the industry is necessary.

---

[24] Smith, H O et al. "Generating a Synthetic Genome by Whole Genome Assembly: phiX174 Bacteriophage from Synthetic Oligonucleotides". *Proceedings of the National Academy of the Sciences* (USA), December 2003. (100) 26.

[25] As noted, in the Sloan Foundation Report, although synthetic genomics techniques are "not yet…general laboratory technique[s] that can be carried out by using a standard protocol manual, many workers are becoming adept at [them]" and they are by no means cutting-edge.

[26] Garfinkel et al, 2006.

[27] Some have speculated that because the viral strains sequenced are primarily from laboratory strains passaged many times outside of a host, the sequences available may not produce a virus as virulent as one obtained from a medical or veterinary sample.

[28] Garfinkel et al., 2006.

Who Should Control Oversight of the Industry?

Community Self-Governance

Many commentators have argued that community-based self-regulation is better-suited to controlling the risks of synthetic genomics. Bügl et al. assert that among other advantages of community self-governance, it is less likely to impede continued scientific advances and commercial growth.[29] Whether they make this claim solely on the basis of the likelihood that self-imposed standards will be less stringent than externally-imposed ones, or to one or more other factors, is, however, unclear. Tucker, for his part, suggests that self-regulation may, in the end, serve as a better guarantor against the misuse of synthetic biology, arguing that "the best line of defense is to ensure the personal integrity and reliability of [the] individuals" within the community.[30] This assertion is echoed by Kwik et al., who hold that "a superior alternative" to government regulation may be "the deliberate creation of an open and expansive research community, which may be better able to respond to crises and better able to keep track of research whether in the university or in the garage", suggesting that "the best way to keep apprised of the activities of both amateurs and professionals is to establish open networks of researchers, perhaps modeled on the Open Source Software (OSS) movement."[31] Under this model, the role of government would potentially be limited to a sponsorship role during the networks' embryonic stages. It is unclear, however, why in particular, these authors believe that the scientific community would more reliably identify rogue or irresponsible research than government agencies.

Further purported advantages of community self-regulation are described by Maurer et al. in the Berkeley report, which argues that self-governance allows scientists to take responsibility for preventing potential misuses of their work, with the implication that this may cause them to consider its repercussions more carefully.[32] The NSABB is currently drafting recommendations for controlling dual-use biological technology. In this report, we further expand upon the recommendations already presented by the NSABB, and include information about the costs, benefits and feasibility of many different options for limiting the risk posed by the unregulated access to custom nucleic acids. They claim, moreover that self-governance mechanisms are generally more quickly and easily implemented than government regulations. While, the authors provide no explicit justifications for these assertions, they are largely in agreement with a broad consensus in the literature. In addition, they argue, self-governance schemes are, due to their essentially consensual nature, more acceptable to regulated communities "and frequently more elegant than externally imposed solutions." Their subsequent (and also largely unsubstantiated) claim that community-based self-governance schemes are intrinsically international in nature is, however, substantially unconvincing. If one imagines the authors' meaning to be that any self-governance system is *necessarily* and instantaneously international in scope, the argument becomes immediately and obviously untenable. It is after all, hardly inconceivable to imagine that adherence to, or even mere formal adoption of, self-governance protocols might be limited to a certain subset (for example, the EU and North America) of the CSNA community, which is, as many authors have rightly noted, truly global in nature. If, alternatively, the authors' intention is to describe a locally- (broadly defined) developed and adopted system of self-governance that subsequently diffuses throughout the global synthetic genomics community and is eventually adopted as an "industry standard", there is no reason to consider self-governance as significantly superior in this regard to government regulation. Governments certainly establish their share of "global industry standards" in any number of fields over which they exercise regulatory authority. Perhaps more convincingly, Maurer et al. go on to argue in some detail that community self-governance could help to promote the emergence of community-wide norms and practices with regard to scientific ethics, biosecurity, and biosafety through means which may not be

[29] Bügl et al., 2007.
[30] Tucker, JB, 2003.
[31] Kwik et al., 2003.
[32] Maurer et al, 2006.

available to, or appropriate for, governments. Ultimately, the proposals contained in the Berkeley report proved too controversial for the synthetic genomics community and were rejected by the attendees of the Synthetic Biology 2.0 meeting held in May of 2006 in Berkeley, California.[33,34] The proposal also drew criticism from outside activists, who alleged that supporters were merely trying to forestall public debate and government regulation and that the report neglected environmental and other risks.[35] However, some attendees have also suggested that the rejection of these proposals may have had more to do with the timing and personalities of their proponents than with their content.[36]

Calls for self-governance from within the CSNA industry, such as the Berkeley proposals often draw comparisons with the 1975 Asilomar conference, at which leaders in the then-embryonic field of genetic engineering agreed to a moratorium on certain categories of experiments and formulated proposed guidelines, subsequently adopted in large part by the National Institutes of Health and promulgated in the first *NIH Guidelines for Research Involving Recombinant DNA Molecules* in the following year. These rules required institutions receiving NIH funding for research involving recombinant DNA technology to establish "Institutional Biosafety Committees (IBCs) to assess the risks of proposed experiments and set requirements for physical and biological containment."[37] More complex issues that cannot be resolved at the institutional level are referred to the NIH Recombinant DNA Advisory Committee. As Jonathan Tucker notes, "this committee has gradually relaxed the NIH guidelines", which are generally agreed to have "placed few obstacles in the path of scientific progress."[38]

Consequently, many proponents of self-governance, including Maurer et al., have self-consciously identified their proposals with the Asilomar protocols, which demonstrated, they argue, "that biological research communities can and do adhere to voluntary standards."[39] Calls for "another Asilomar" to address synthetic genomics typify these sentiments. But, as the authors of the Sloan Foundation Report attest, "Asilomar is for the most part not the right model" for regulating the synthetic biology community. The protocols established there, they note, were first and foremost, "an exercise in *self*-governance" undertaken to assuage *safety* concerns both within and outside of the scientific community. "Although there was apparently some discussion over whether to consider the possibility of biological warfare [or not], it was decided that the hazards of carelessness were enough to deal with at that meeting." Never intended to confront the challenges of even state-sponsored bioweapons, they argue, the Asilomar model of self-governance is dangerously inadequate in the face of bioterrorism. "Bioterrorists", they observe, are "by definition…[unwilling] to accept the norms of the research community, and no community can control all subsequent uses to which the research results or techniques it develops might be put." Thus, they conclude, self-governance must be combined with other, mandatory measures to adequately confront the potential threat of bioterrorism. Furthermore, as Tucker points out, the current Asilomar protocol-inspired *NIH Guidelines* place no obligation upon private companies and institutions that receive no NIH funding. An effective "system for the security oversight of dual-use research in the life sciences", including of course, synthetic genomics, "should ideally go beyond the scope of the NIH guidelines by covering three types of activity: (1) government-funded or privately-funded academic studies conducted with the expectation of publication; (2) proprietary research by private industry that is not intended for

---

[33] Aldhous, P. "Synthetic Biologists Reject Controversial Guidelines", *New Scientist*, May 23, 2006.

[34] According to the *New Scientist* story, some critics of the proposals "argued that it [was] too early to boycott gene synthesis firms, as it is not yet clear how best to screen for sequences that might be used to make a bioweapon." They further object that currently, "there are…no clear channels through which dangerous experiments could be reported." The meeting declaration adopted by the convention instead pledged "to help develop software and other tools to improve companies' ability to identify orders for potentially dangerous DNA." *Ibid.*

[35] Aldhous, P, 2006.

[36] Private correspondence with Synthetic Biology 2.0 attendee who wished to remain anonymous.

[37] Tucker, 2006.

[38] Tucker, 2006.

[39] Maurer et al., 2006.

publication; and (3) classified government research conducted for purposes of threat assessment or biodefense."[40]

Government-Mandated Oversight

While few outside of the anti-GMO (genetically modified organism) activist community argued strenuously for very strict government regulation of the synthetic genomics community, authors with government or policy backgrounds (the NSABB Working Group on Synthetic Genomics and Dr. Jonathan Tucker, for example) tended to assume that a certain level of government regulation of synthetic genomics would be necessary. However, these commentators often tempered their support for such regulations with reminders of the benefits of technology as well as the risks of over-regulation. In particular, these commentators tended to regard community self-regulation as potentially inadequate (as noted above) in the face of biosecurity, as opposed to biosafety threats. On the other hand, authors purporting to represent the synthetic genomics community, such as Maurer et al., tended to support community self-regulation and to fear that government regulation might impede scientific discovery and commercial development.

The primary advantages of a system of government regulation (though they are generally expressed as means of overcoming the shortcomings of community self-governance) are, as the NSABB Working Group on Synthetic Genomics notes: its uniformity (at least within any given jurisdiction) and the ability of the regulating body to effectively enforce its regulations. In the particular case of proposed regulations for synthetic genomics, the NSABB Working Group concludes that uniform and standardized screening and data storage practices among commercial DNA synthesizers are needed to safeguard against intentional or unintentional circumventions of existing regulations, such as the Select Agent Rules. If one accepts the Working Group's further stipulation, that "effective compliance" with said standards "may also require audits, fines and/or other legal actions", then government regulation may be the best or only way to minimize security risks. The NSABB recommendations also raise another putative advantage of government regulation – the far greater ability of the government to ensure compliance, through inspections, audits, and other forms of monitoring, with a given body of regulations and to penalize those who fail to do so. Nonetheless, in our view, regulatory uniformity may prove a decidedly mixed blessing. As the NSABB Working Group's own conclusions regarding controls placed on research with variola (smallpox) virus illustrate, more discriminating regulatory standards are often desirable. Furthermore, the superior enforcement powers available to government regulators do not guarantee their use, as in the enforcement of Commerce Department Regulations, as discussed further in Chapter 4.

Aside from the *NIH* Guidelines, perhaps the most relevant US Government regulations on dual-use biological research promulgated in the past three decades are the Select Agent Rules, introduced in 2002. As Ronald Atlas notes, most among the scientific community have come to see the Select Agent Rules as part of "a number of sound steps" taken by the government in the aftermath of 9/11 and the subsequent anthrax attacks "to prevent the acquisition by terrorists of dangerous pathogens and toxins that could be used against civilian populations." Despite fears "among scientists that the national security concerns will undermine scientific progress," the government has largely eschewed a "heavy-handed approach."[41] Nonetheless, such fears and attendant frustrations with the Rules persist among some in the scientific community.

However, the shortcomings of government regulation are legion and critics are hardly slow to point them out. As the authors of the Sloan Foundation Report note, any government-mandated regulations of

---

[40] Tucker, 2006.

[41] Atlas, Ronald M. "Securing Life Sciences Research in an Age of Terrorism", *Issues in Science and Technology*, September 2006.

synthetic biology are likely to be purely national in nature and consequently may have little impact on synthesis abroad. As the National Research Council Report, *Biotechnology in an Age of Terrorism* observed, "any serious attempt to reduce the risks associated with biotechnology must ultimately be international in scope, because the technologies that could be misused are available and being developed throughout the globe."[42] Certainly, synthetic genomics is no exception. Some proponents of government regulation, however, have argued that it is only natural for the United States, given its leading position in the field of synthetic genomics, as well as biotechnology more generally, to take the lead as well in controlling its potential security and safety risks. Moreover, as the NSABB Report on Synthetic Genomics reasons, the establishment of a domestic system of regulation can be pursued in parallel with a US Government effort to "foster an international dialogue and collaboration" aimed at "developing and implementing universal standards and preferred practices for screening sequences" and other security concerns related to synthetic genomics. Indeed, one might quite reasonably regard the establishment of domestic and global standards on screening and other security measures as a sort of mutually-reinforcing positive feedback loop. For example, the recent adoption by the UK of a visa-screening process (entitled The Academic Technology Approval Scheme) for foreign (non-EU) students wishing to pursue graduate study in certain "sensitive" areas follows the example of the similar US system first introduced in 2003 and administered by the Interagency Panel on Advanced Science and Security.[43,44]

A further objection to top-down government regulation of biological research is raised by Kwik et al., who note that such directives often run "the risk of being either too heavy-handed and all-encompassing, thereby interfering with critical life-saving research, or of missing work that is vulnerable to [misuse]…and applicable to potential bioweapon applications."[45] Of course, it should be noted, that the authors' objections could, at least in theory, just as easily be made to any other system of regulation, whether consensual or non-consensual, government-mandated, or not. Many commentators have argued, however, that community-based self-governance may prove more flexible and adaptable to changing scientific, commercial, and security circumstances than government statutes and regulations. The latter, Kwik et al. argue, may often "impose 'one-size-fits-all' constraints", under which "the 'fit' between the hazards subject to regulation and the actual circumstances of application is often imperfect. Enormous effort and expense may [consequently] be lavished on relatively small risks, while more important issues are neglected."

Many commentators, including Kwik et al. have also raised the concern of cost with regard to government regulation of CSNA, or biotechnology more generally. They contend that the costs of government regulation of biotechnology can be extensive, generally exceed those of community self-regulation, and "given the fierce competition for research funds and the uncertain payback of most R&D efforts, [are] possibly unsustainable."[46] Another problem that has arisen in the implementation of the Select Agent Rules, as Tucker notes, is their imposition, in some cases, of "an undue financial burden on academic researchers, especially those working in smaller laboratories."[47]

The legacy of government regulation of dual-use biological research is, moreover, by no means one of unalloyed success. As Benner and Sismour caution, who cite the ban by the City of Cambridge, MA on

---

[42] National Research Council, *Biotechnology Research in an Age of Terrorism*, 2004.
[43] Dyer, NJ. "Unis [sic] Must ID Students in 'Suspicious' Subjects", *The Times of London*. March 20, 2007.
[44] From the Testimony of Janice Jacobs, Deputy Assistant Secretary of State for Visa Services before the Committee on Science, US House of Representatives, March 26, 2003. "Foreign Students and Scholars in the Age of Terrorism". Available at: http://www.state.gov/r/pa/ei/othertstmy/33002.htm
[45] Kwik et al., 2003.
[46] Kwik et al., 2003.
[47] Tucker, 2003.

recombinant DNA research, the potential benefits of any technology must always be considered alongside its hazards, something which regulators have not always been in a position to do.[48]

However, as the authors of the Sloan Foundation Report argue, even in the event that policymakers decide that current regulations are sufficient to accommodate the challenges of synthetic biology, "guidance is" nonetheless "needed merely to apply existing regulation or governance mechanisms, defined in the context of one set of technologies, to a subsequent technological generation."[49] In particular, as noted by the NSABB Working Group on Synthetic Genomics, the US Government may find it necessary to consider "whether genomic material that does not exactly match the genomes referenced" in the Select Agent Rules, should be subject to these regulations.[50] We note that because of the natural variation within strains of viruses and species of bacteria, these mismatches could be rather common. Moreover, policymakers must confront the possibility that ongoing technological developments may soon render current regulations obsolete, or irrelevant. As CSNA techniques and technologies become more widespread, circumvention of the Select Agent Rules (or analogous regulations) through the synthesis of listed agents or toxins will become increasingly likely. Governments may find themselves compelled to issue further regulations merely to guard against the subversion of existing laws and regulations.

Requirements, Needs, and Limitations of an Oversight Regime

Much has been written about the requirements and needs that must be taken into account when proposing regulation of the oligo synthesis industry. Many of these requirements focus on the need for synthetic DNA regulations to reach globally, to impose minimal costs both economically and socially on the industry while fitting into a larger regulatory picture and creating a framework for reporting suspicious activity.

Global Action

One requirement that has been stressed in nearly every article on the subject of synthetic biology is the need to think globally. Many feel that any regulations that do not encompass the entire, global, CSNA industry will cause undue harm to the both the regulated companies that produce CSNA and the scientists that rely on these products, without significantly increasing biosecurity. Owing to this fact, many point to the serious limitations that must be addressed when contemplating regulation of the CSNA industry.

Finding common ground on the issue of CSNA within a single country or even a small group of individuals has proven difficult. Probably due to the devastating and shocking events on 9/11, the risk of bioterrorist attacks seems to be experienced differently in the US than in Europe as pointed out by Vriend in the report "Constructing Life: Early social reflections on the emerging field of synthetic biology,"[51] further complicating the global solution. While the Swiss Genetic Research Forum has adopted an ambivalent position on the issue of biosecurity, arguing that the possibility of the abusive and criminal application of synthetic biology, is negligible,[52] members of the US synthetic biology community have called for greater vigilance against biosecurity threats.[53] Reflecting upon this divergence in views, Chyba and Greninger express unease in their report "Biotechnology and Bioterrorism: an Unprecedented World,"[54] that a 'crazy-quilt' pattern of oversight would likely vary from country to country, and could be

---

[48] Benner and Sismour, 2003.
[49] Garfinkel et al, 2006.
[50] NSABB Working Group on Synthetic Genomics, 2006.
[51] De Vriend, H. Constructing Life: Early social reflections on the emerging field of synthetic biology. The Hague: the Rathenau Institute, 2006.
[52] De Vriend, 2006
[53] Maurer et al, 2006.
[54] Chyba, CF and Greninger, AL. "Biotechnology and Bioterrorism: an Unprecedented World." *Survival* **46**(2), 2004.

non-existent in countries of greatest concern. We would, however, be remiss were we to overstate the perceived divergence in American and European attitudes towards biosecurity issues. As discussed further in Chapter 4, the European Commission is currently considering tough new measures designed to address the risks inherent in dual-use biotechnology, including plans for an EU-wide system of security clearances for researchers, institutions, and experiments as well as a proposal for restricting the publication of "sensitive research findings."[55] In a further sign of growing European biosecurity awareness, another measure under consideration would mandate that all European undergraduate science students "take lessons in ethics to raise awareness of the ways in which their work could be exploited by terrorists."[56]

Nonetheless, even given the ongoing trans-Atlantic convergence of attitudes toward biosecurity issues, regulatory consistency remains an important concern. Indeed, as Laurie Zoloth of the Center for Bioethics, Science, and Society has argued, regulatory oversight of "dual use" research in molecular biology has not always been consistent even within the United States.[57] Moreover, the pace of progress in the development and diffusion of biological technologies is notoriously swift, leading Robert Carlson[58] to conclude that "It is unrealistic to think biological technologies can be isolated within the boundaries of officially sanctioned countries. Even if such a regime were implemented, it would merely include those countries that already have a particular technology. We can do little to take technology away from those in whose hands it was developed and resides."

Screening as Part of a Larger Oversight Scheme

In the article "Biotechnology and the Challenge to Arms Control"[59] Chyba points out that synthetic DNA does not necessarily need to be purchased commercially. Although beyond the scope of our study, Chyba's point is an important one; any regulations of CSNA providers must take into account that oligo and synthetic gene manufactures do not provide the only sources of CSNA; to prevent unauthorized access to CSNA, access to DNA synthesizers, chemicals, and even pathogenic sequences must be considered.

Harvard Biologist George Church suggested that all new DNA synthesis machines manufactured be licensed, tagged with electronic locators, and programmed to forbid the synthesis of dangerous DNA sequences. But according to Robert Carlson in the article "The Pace and Proliferation of Biological Technologies",[60] "Even if great care is taken to limit the commercial synthesis of DNA from pathogens or toxins, it is unlikely the chemical tricks and instrumentation that companies develop in the course of building their businesses will remain confined within their walls. Eventually, efficient synthesis will be possible using instruments assembled at home." In fact with a quick search on Ebay or LabX.com[61] it is easy to pull up hundreds of DNA synthesizers for sale, many at less than $10,000. With the sale of used scientific equipment on the internet having become commonplace in recent years, it is difficult to imagine that all synthesizers currently in laboratories could be easily tracked down.

---

[55] Charter, D. "University Researchers to Be Vetted to Tackle Bio-Terror Threat", David Charter, *The Times of London*, July 10 2007. Available at: http://www.timesonline.co.uk/tol/news/world/europe/article2056570.ece

[56] Charter, 2007.

[57] Zoloth, L, "Ethical Issues in Synthetic Biology. Security and Regulation of Experiments of Concern. A White Paper on the Ethics of Self-Governance in New Scientific Community", found as an appendix to Maurer, et al 2006.

[58] Carlson, Robert. "The Pace and Proliferation of Biological Technologies". *Biosecurity and Bioterrorism: Biodefense Strategy, Practice and Science*. September 2003. (1) 3.

[59] Chyba CF. "Biotechnology and the challenge to arms control" *Arms Control Today*, October 2006.

[60] Carlson, 2003.

[61] LabX.com was founded in 1995 to provide forum where buyers and sellers of new, used, surplus, and refurbished scientific and laboratory equipment could find items, negotiate terms, and complete transactions online.

We note, though, that the trend that Carlson highlights in his article is driven by the purchase of ever more complicated and expensive capital equipment by a few dedicated oligo synthesis companies, not a growing capability of instruments suitable for any laboratory's use. This fact suggests that the capabilities of those synthesizing DNA without ordering oligos from dedicated suppliers may, in fact, not be growing appreciably and brings into question the assertion that in the near future most researchers will produce their own oligos.

Costs to Industry, Researchers, and Government

A second need pointed out in the Berkeley Report[62] that should be addressed by any regulation of this industry is the need to keep costs low. The Berkeley Report values the benefits of synthetic biology in the tens of billions of dollars and states that "synthetic biology stands to generate still larger gains by creating products that cannot be achieved by traditional methods… Any intervention that threatens these developments is likely to be counterproductive." Although the Berkeley Report is concerned with more than just the CSNA industry, it should be noted that the CSNA industry underpins much of the synthetic biology community. Therefore, destructive regulations imposed upon the CSNA industry would likely also cripple the synthetic biology community, greatly endangering the advances in biomedicine that may flow from its research.

According to Bügl[63] et al., in the past 10 years oligonucleotide costs have fallen by a factor of 10 to $.20/nt and expected delivery times are currently ~48 hours (from order to delivery). These prices are slightly lower than those reported by BioInformatics[64] (who surveyed the marketplace) as well as by examination. By comparison today's synthetic gene prices average $1.00 per nucleotide, when desired modifications and purification are averaged in. With these cost points in mind, Bügl et al. believe the best framework would integrate regulation into the normal commercial synthesis operation at a modest cost without having a major impact on the speed at which the customer receives the product. Given the low cost of current products and the rapidity at which the products are delivered, the previous statement by these authors must be interpreted to mean that the system overall can impose a very minimal cost burden on the products. Interestingly, the authors of the Berkeley Report suggest that "most gene synthesis companies already screen" suggesting that costs would not rise among these companies. In contrast, as discussed further in Chapter 3, although we found that most gene companies do screen orders, we have found that most companies do NOT screen orders for oligos, beyond confirming that the purchaser has the ability to pay. Given that oligos comprise the bulk of the volume of the CSNA industry's orders, the statement in the Berkeley Report should *not*, in our view, be interpreted to mean that the cost of adopting screening procedures for the majority of the CSNA industry is likely to be minimal.

Bugl et al. go on to say that any regulation of the industry should maintain the ability to deliver high-quality products at low costs with very rapid delivery times. This brings up the important point that not all cost associated with regulation can be measured in strict monetary terms. Many companies currently offer next day delivery on CSNA orders. Regulations that would slow the shipment of these orders would cause hardship for both the customers who rely on CSNA, and the companies that sell these products. This point brings us back to the importance of global regulation. Bugl et al.[11] note that in order to ensure a fair playing field for all companies in the international CSNA market, any regulations on the industry should not hinder a single country or group of countries' commercial market without international consensus.

---

[62] "Community-Based Options for Improving Safety and Security in Synthetic Biology"-Berkeley Report.

[63] Bugl, H. et al. "DNA Synthesis and Biological Security." Nature Biotechnology. 2007.

[64] BioInformatics. "The Global Market for Synthetic Oligonucleotides." Report #06-058, September 2006.

Each potential option for regulation, according to the Sloan report "strikes a different balance between concerns for the potential harm that might be posed by synthetic genomics versus concern about foregoing synthesis' benefits or about imposing other costs on society."[65] However in the article "Limiting Terrorist Access to Deadly Pathogens"[66] J.B. Tucker argues that some regulations may have a chilling effect on legitimate research, pointing to a sudden decrease in research on Select agents at Duke University when six of 57 laboratories working on Select Agents discontinued their work after regulations went into effect. Although this figure may seem significant, we are unsure how often laboratories switch focus in the absence of new regulations. Given that the Select Agent regulations are very strict (as detailed in Chapter 4) and that regulations of CSNA will likely be less stringent, a similar decrease in CSNA-based research will most likely not been seen, even if this decrease is directly attributable to regulation. Because of the vast array of benefits from CSNA, many authors stress the importance of choosing regulations that will not impede legitimate research.

A Framework for Reporting

As you will read in Chapter 3, screening methods have been developed to prevent hostile actors from ordering CSNA sequences that could potentially be used to create Select Agents and many gene synthesis companies are already screening for just that. When considering any regulation that involves screening of customers or sequences one needs to consider the actions that should be taken if a screening system is to recognize a violation, and how information will be stored for forensic use in the event of a screening system failure.

During the course of our industry interviews many companies noted that they weren't aware of a government contact to notify in the event of a potential problem. Vriend suggests the creation of a confidential hotline for biosafety and biosecurity issues,[67] while Wheelis suggests reporting to the Department of Homeland Security, the Centers for Disease Control or the FBI.[68] While there is not a consensus on the most appropriate contact agency, from the literature available on the subject it appears that there is a need to address the issue of how to handle those orders or customers that are identified as suspect. However, many of our industry interviewees worry that if the criteria for reporting are not stringent enough, a hotline such as that suggested by Vriend would be flooded with calls rendering it virtually useless.

In addition to providing a government contact so that those in industry can report suspicious orders, there is also a need to develop a framework for investigating completed orders that escaped detection. In his letter to the Editor of *Genome Technology*,[69] John Mulligan, Chairman of leading gene synthesizer Blue Heron, points out that measures to prevent the misuse of CSNA for bioterrorism cannot eliminate all risk. The Sloan Foundation Report[70] touches on the need to maintain records such that those who sell CSNA would be able to help with forensic investigation if CSNA were used by hostile actors in a biologicial attack. In other words, these authors claim it is important that CSNA suppliers archive order information. Similarly, the NSABB Working Group on Synthetic Genomics[71] recommended that the US Government specifically "develop standards and practices to be used by [CSNA] providers for retaining records of orders for gene-length or genome-length nucleic acids, and require that records be retained by

---

65 Garfinkel et al, 2006.
66 Tucker J.B. Biosecurity: Limiting Terorrist Access to Deadly Pathogens. 2003. United States Institute of Peace. Washington, D.C.
67 De Vriend, 2006.
68 Wheelis, M. "Will the New Biology Lead to New Weapons," *Arms Control Today*. July/Aug. 2004.
69 Mulligan, J. "Letter to the Editor." *Genome Technology*. March, 2003.
70 Garfinkel, et al, 2006.
71 NSABB Working Group on Synthetic Genomics, 2006.

providers… If there is to be a review or use of the results of [a] screening effort", they argue, "records will need to be retained." The authors furthermore endorse the establishment of a standardized protocol for record-keeping as "the least intrusive way to accomplish effective implementation and compliance", which they imagine may need to be ensured by the means of "audits, fines and/or other legal actions" by the government.

When considering regulation of the CSNA industry we must look at the requirements and needs that should be taken into account, as well as the limitations that must be considered and overcome. As noted above, the literature stresses that regulation of the DNA industry must consider the global perspective, keeping in mind both manufacturers outside of the United States as well as individuals who choose to synthesize their own DNA. They must be conscious of the benefits of the technology and make an effort to keep costs (both economic and social) to a minimum. Lastly, regulations should be constructed within a framework that makes reporting suspicious activity straightforward while allowing easy access to archived order information in the event of misuse of the technology. The next section details regulations suggested in the literature and how each attempts to deal with the needs and limitations presented above.

Policy for Sequence-based Screening

The Sloan Foundation[72] details a plan that would require CSNA providers to use approved software to screen orders for pathogenic sequences. This is echoed by the NSABB[73] Working Group on Synthetic Genomics, who suggest that funding agencies, in particular the National Institutes of Health, mandate that government grantees and contractors order only from CSNA suppliers who employ certain safety and security precautions, including sequence screening with government-approved software. Although this kind of software is already commercially available under the name BlackWatch marketed by Craic Computing and many in the gene synthesis industry have developed their own software to perform this function, the Sloan report calls for better software as well as a well-designed and constantly-updated list of sequences against which orders will be screened. The Berkeley Report[74] agrees that sequence screening by synthetic gene providers is important and advocates that the synthetic biology community create community-wide standards insisting that all commercial gene synthesis houses adopt current best practice screening procedures. Neither group feels that there would be a significant burden associated with screening, however, once again it should be stressed that these groups were probably considering only the screening of genes, not the screening of oligos. They go on to propose that members of the scientific community boycott companies that do not screen all orders. However, this proposition drew considerable criticism from within the CSNA community itself, with many scientists arguing that it was "too early to boycott gene synthesis firms, as it is not yet clear how best to screen for sequences that might be used to make a bioweapon", and contributing in large part to the rejection of the Berkeley Report proposals at the Synthetic Biology 2.0 conference.[75] Moreover, given the non-uniform[76] nature of current industry screening practices detailed in Chapter 3, a community boycott may be difficult to achieve. While many synthetic gene companies are already screening sequences for pathogen containing elements, others screen only rarely or not at all, and sequence screening is virtually non-existent in the oligo industry.

Moreover, from a technical standpoint, current screening software is far from ideal. The Sloan report points out that the first generation software available today has difficulty identifying the organism of

---

[72] Garfinkel, et al, 2006

[73] NSABB Working Group on Synthetic Genomics, 2006.

[74] Maurer, et al, 2006.

[75] Aldhous, 2006.

[76] Some companies report screening all orders, others screen only a portion or none at all. Screening methods also vary from commercially available software, to internally developed software, to a simple BLAST screen as described further in Chapter 3.

origin for short (oligo length) pieces of DNA and add "since oligos are used in a wide variety of different applications, the sheer volume of production of oligos far exceeds that for synthesis of genes and genomes." Therefore sequence screening may prove more difficult for oligo length sequences given currently available screening software. Sequence screening is further complicated for oligo providers by the high throughput nature of the oligo synthesis business. Under current best practices, sequences identified as potentially containing a pathogenic element of concern must be manually examined; this is impractical for companies that are receiving thousands or tens of thousands of orders a day.

Another shortcoming of existing screening software is that it may not include all potentially dangerous sequences. Firstly, databases currently used in conjunction with screening software may only include select agents and toxins, leaving out potentially dangerous sequences from other pathogens. Secondly, as noted by the NSABB Working Group on Synthetic Genomics, "it is now feasible to produce synthetic genomes that encode novel and taxonomically unclassified agents with properties equivalent to, or potentially more harmful than, current Select Agents."[77] As a result, current screening practices, which focus on sequence homology with known pathogens, may be insufficient to detect all potentially hazardous sequences. On an even more fundamental level, it remains controversial what sequences are potentially hazardous. Due to the concerns described above, many commentators including the authors of the Berkeley report[78] have concluded that new screening software must be developed and a database of potentially hazardous sequences must be built to support the software.

Current screening practices may be unable to detect dangerous sequences that have been split into multiple orders, perhaps obtained from multiple providers. Therefore, Maurer et al.[79] endorse surveillance across multiple orders. A centralized screening facility could ameliorate these concerns, but according to our industry contacts may raise concerns surrounding confidentiality and thereby encourage customers to synthesize in-house (particularly the pharmaceutical industry).

What Should be Screened?

When considering the implementation of sequence screening one must assess whether some orders should be exempted from screening and if so on what basis. Screening all orders may lead to an unacceptable cost burden from the screening itself or high false hit rate – the shorter the DNA sequence, the lower the confidence that the particular sequence of concern is found exclusively in a pathogenic organism (let alone that it is associated with pathogenicity).[80] Moreover, oligo-length DNA sequences comprise the bulk of CSNA orders. Thus, the burden of screening oligos may be disproportionately large in relation to the biosecurity risk reduction.

Due to the increased difficulty associated with constructing a pathogen from oligos (rather than longer gene-length pieces of DNA) and the high number of database matches that would likely result from screening oligos, some have proposed that "screening could be required for only longer sequences (for example, greater than 500bp) or for all commercially synthesized DNA."[81] While these are certainly important caveats to bear in mind when considering a proposed sequence screening protocol, the fact remains that the ability to construct a genome from much shorter oligos is a proven fact (as evidenced by the success of both the Wimmer group with poliovirus and the Venter Institute group with phi X 174) and may be becoming progressively easier as technology and synthesis techniques continue to develop. The authors of the Sloan Foundation Report may also have failed to account for another potential risk posed

[77] NSABB Working Group on Synthetic Genomics, 2006.
[78] Maurer, et al, 2006.
[79] Maurer, et al, 2006.
[80] Garfinkel, et al, 2006 and Tucker, 2006.
[81] Garfinkel, et al, 2006

by unfettered access to synthetic oligos – the amplification via PCR (polymerase chain reaction) of the genes encoding dangerous toxins or other pathogenic elements.

Strategies for Use in Conjunction with Sequence Screening Models

One proposed strategy that could be used alone or in conjunction with sequence screening is the licensing of scientists and/or labs. The Sloan Foundation[82] presents an option to replace requirements for industry based screening with an institution-based certification of legitimate researchers. "Under this option staff that place orders for synthetic DNA must be first certified as legitimate users by that institution's biosafety officer. In order to accept an order, commercial firms would need to see that the individual researchers had been approved to place such orders by a registered institutional biosafety officer." Rather than having to screen each individual shipment, the biosafety officer would certify as to the legitimacy of each user. To keep certification current, users may be subject to biosafety training requirements. In addition the biosafety office may choose to re-examine an individual's research, or individual orders.[83] A centralized and continually updated, electronic list of certified researchers would be made available to synthetic DNA providers "so that individual orders could be approved with minimal time delay." Garfinkel et al. go on to note that such a system could be modeled after the approach used by the American Type Culture Collection (ATCC), which we discuss further in Chapter 4.[84]

Bügl et al.[85] propose that institution-based certification be used in conjunction with sequence screening. The authors of the Sloan Foundation Report[86] considered this option as well. "The biosafety official would be asked to certify researchers into two categories: 1) legitimate users of synthetic DNA, and 2) researchers with ongoing experiments with pathogens or with DNA that might come from a pathogen." If a sequence ordered by a researcher (working in a US lab regardless of nationality) in the first category is identified as potentially hazardous, the institutional biosafety officer would have to be consulted for approval before the order could be filled. Researchers in second category would be exempt from this verification process. Furthermore, in an analogous manner to current radionuclide ordering procedures, "shipments of certain types of hazardous genes or portions of pathogens, instead of being shipped directly to the individual researcher, might be sent to the institutional biosafety officer or to the chair of the Institutional Biosafety Committee" (IBC).[87] George Church[88] proposes that researchers certified to order potentially hazardous sequences could be exempted from sequence screening.

The institutional based certification proposals as acknowledged by some proponents, fail to address certain key biosecurity concerns. For instance, not all orders for CSNA originate from institutions that are required to abide by the biosafety procedures established in the NIH guidelines. Furthermore, as the authors of the Berkeley Report note, there is doubt in the community that institutional biosafety officers and IBCs "have sufficient biosecurity expertise to screen experiments of concern." More generally, as previously discussed, current biosafety practices may not adequately address biosecurity concerns. In particular, as Chopra and Kamma[89] note, institutional licensing, on its own, does little to prevent the malicious synthesis of a pathogen or toxin. For instance, there seems to be few checks against a malicious actor establishing a sham biotechnology company and naming himself the biosafety officer. Without inspections to ensure that licenses are legitimately obtained, this system seems to have few teeth.

---

82 Garfinkel, et al, 2006
83 Garfinkel et al., note that examining individual orders would likely slow down the approval processes.
84 ATCC will only ship potentially hazardous material with the approval of a registered biosafety professional.
85 Bügl et al, 2007.
86 Garfinkel, et al, 2006.
87 Garfinkel, et al., 2006.
88 Maurer et al. 2006.
89 Chopra, P and Kamma, A. "Engineering Life through Synthetic Biology" *In Silico Biology*, **6**, 2006.

Furthermore, a reasonable practice for licensing and verifying the license of foreign customers of US CSNA suppliers has yet to be described.

A third option suggested in the Berkeley report involves the use of special DNA "bar codes" that could be inserted into CSNA for facilitating detection and deterrence, as well as authorship and responsibility. "Sequences could be optimized so that organisms containing selected bar codes could be readily identified using PCR" whether for the purposes of screening or forensic investigation. Said sequences may serve to deter hostile actors and may foster a feeling of responsibility and authorship while facilitating the enforcement of intellectual property rights. Despite their perceived benefits, genetic bar codes face a number of difficulties that must be resolved prior to implementation. Firstly, as Maurer et al. allow, "despite preliminary experiments, bar coding technologies have yet to be demonstrated. In particular; it is not clear whether bar codes would be stable against mutation," or whether they can be readily detected and removed. Secondly, "bar codes would inevitably interfere with experiments involving short DNA sequences." Lastly, in our view, bar codes only make sense when an organism is being created. If the tagged oligos are used to PCR-amplify a target gene (that encodes a pathogenic element, for instance) for inclusion into a non-pathogenic bacteria, the bar code is likely to be excised in the process of cloning the gene from the PCR-amplified fragment.

Yet another regulatory option, as proposed by Gesteland et al.,[90] involves "monitoring DNA sequences shipped from DNA synthesis facilities capable of producing large segments of DNA…much like the Drug Enforcement Agency attempts to monitor chemical purchases to detect drug-making activity." However, as our industry interviews indicate, the quantity of clients and shipments of the synthetic gene industry may make government monitoring of all shipments originating from synthetic gene providers nearly impossible.

All discussions on oversight of the industry should consider that CSNA is not intrinsically dangerous. This is due to the fact that only when the CSNA encode sequences from a pathogen and is used to construct something harmful does it become dangerous. With this in mind, some authors propose the regulation of the information that could lead to the ordering of "dangerous sequences". The recreation of the 1918 flu virus stirred controversy in 2005, as some "hailed the work as giving unprecedented insight into the virus."[91] "Working out how it arose and why it was so deadly", they argued, "could help experts to spot the next pandemic strain and to design appropriate drugs and vaccines in time."[92] Critics however, charged that the publication of the research findings (including the full genome sequence) was dangerous and irresponsible. As *Nature* reported at the time, citing an unnamed "biosecurity expert",[93] "the publication of the full genome sequence gives any rogue nation or bioterrorist group all the information they need to make their own version of the virus." With such concerns in mind, Chyba and Greninger[94] propose that the synthetic biology community adopt a form of publication review under which scientists and editors of scientific journals would review their own publications for potential national security risks. In some cases, certain information (such as potentially dangerous sequences) might be withheld from publication. However, as noted by Atlas[95] there is no consensus on what information should be withheld from publication which may lead to an inconstant system of rules and regulations. Such a system is, in our view, unlikely to address many biosecurity concerns raised by the publication of the 1918 flu virus genome sequence. After all, the researchers received permission to conduct the study from both the CDC and National Institute of Allergy and Infectious Diseases (NIAID) Directors and the publication of the findings was approved not only by the editors of *Nature*, but by an emergency session of the NSABB as

[90] Gesteland et al, 2004.
[91] Von Bubnoff, A. "The 1918 Flu Virus is Resurrected", *Nature*, October 5, 2006.
[92] Von Bubnoff, 2006.
[93] Von Bubnoff, 2006.
[94] Chyba and Greninger, 2004.
[95] Atlas, 2006.

well, which recommended only that "the authors…add a passage to the manuscripts stating that the work is important for public health and was conducted safely."[96]

Nonetheless, it is the consensus in the scientific community that the advancement of scientific inquiry requires an open system of information sharing. It may become more difficult for other researchers to verify and reproduce a given author's experimental results under any proposed system that limits the exchange of scientific information. As Jeffery Taubenberger, lead author of the 1918 flu sequencing study told *Nature*, there can be no absolute guarantee of safety, whether from biological accidents or attacks. "We are aware that all technological advances could be misused", he said. "But what we are trying to understand is what happened in nature and how to prevent another pandemic." The NSABB is currently drafting a proposed framework for the oversight of dual-use life science research that contains strategies to minimize the potential misuse of research information. Their draft document will address recommendations on dual use communication among other issues.

Conclusions

There is an ongoing dispute within the literature as to whether or not the synthetic DNA industry should be regulated. While some feel that synthetic DNA poses a great biosecurity risk and therefore regulation is necessary, others hold that the risk of misuse is low and that regulation will inflict more harm than good. Those calling for regulation are conflicted about who should be in charge and what methods should be used. Most who have written about this subject recognize that there are many issues such as social and economic costs and global implications that must be taken into account when proposing regulation of the industry. Many who support regulation feel synthetic DNA sequences should be screened and/or customers licensed before products are sold. Others call for tagging sequences and detailed record-keeping. It is clear from the literature available that there is no silver bullet that can prevent misuse of the technology and that any regulation imposed will have a set of costs associated with it. This being said, one must also keep in mind the costs associated with maintaining the status quo.

Clearly, the lack of consensus on several key issues related to the oversight of the synthetic DNA industry underscores the need for further study. In the chapters that follow, we provide data on the nature of the industry, its customers and its products to help inform decisions on oversight. Finally, we outline options for a notional oversight regime, and highlight the costs and benefits of each option, so that policymakers can arrive at the best course of action.

[96] Von Bubnoff, 2006.

Chapter 3: Industry Review

The Products

There are many types of CSNA products made today for a variety of applications. These include genes, modified oligos, non-DNA-based oligos, oligos of various sizes, and oligos produced in different formats. In this section, we briefly discuss the various products made by the CSNA industry and how these products may impact an oversight regime.

Oligo Length

As described in the Scope section in the Introduction, there are activities of concern that require CSNA as a critical reagent. The activity most often discussed in this context is the de novo synthesis of viruses by the fabrication of their genomes by the concatenation of synthetic genes or oligos. More ambitiously, and further into the future, is the transmutation of one closely related bacterial species into another, perhaps more pathogenic bacteria, by the transplantation of its genome (as demonstrated recently by the Venter group).[97] Less ambitious is the cloning of virulence factors into less pathogenic strains with the hopes of enhancing their pathogenicity. This approach combines synthetic genomics with traditional recombinant DNA engineering approaches and can be accomplished in the absence of a sample that contains the pathogen by ordering the desired gene (or making the gene synthetically from oligos) and inserting this gene into another organism. If a sample that contains the pathogen is available, the desired gene or genes can be acquired through PCR-based amplification using two oligos.

If we look at a theoretical example of someone wishing to synthesize a virus, there are four salient methods by which this could be done. First, the entire sequence of the virus could be ordered directly from a gene synthesis company. Secondly, long, gene-length fragments of the viral genome could be ordered. Depending on the virus, these fragments could number as few as three to as many as several hundred segments if each segment is approximately one gene. Thirdly, the viral genome could be could be ordered in many small overlapping oligonucleotide segments of 35-65nt. And finally, the researcher could PCR amplify the segments of the viral genome from an environmental or medical sample (note that the researcher using PCR must have a sample containing the virus, so this technique is not readily applicable to viruses that are difficult to find in nature.) The researcher would then assemble the viral genome from the pieces ordered or amplified using commonly available recombinant DNA techniques, such as hybridization and ligation (and, for RNA viruses, transcription). Obtaining useful, infectious virus normally involves the introduction of the viral genome into cells that could produce the virus, an additional non-trivial step that will involve manipulations unique to the type of virus to be produced.

In order to perform a trade-off analysis on the cost and benefit of including oligos of various lengths in a screening system, it was necessary to examine what aspects of CSNA could be considered enabling to a hostile actor who wished to construct a pathogen from oligos. Given that viral genomes range from several kilobases to several hundred kilobases in length, and bacterial genomes are much longer, we assume that there is no practical *upper* limit for CSNA size in this system. That is, there is no length of CSNA that prevents its use in a malicious experiment simply by being too long. Therefore, we set out to determine what the smallest CSNA pieces would enable experiments of concern.

PCR can be performed with very short oligos, such that very few oligos are ordered that are too small to be used for PCR. Therefore, excluding the oligos that are too short for PCR would provide very little benefit, if any. Said another way, if the system is designed to detect the malicious use of CSNA in the

[97] Lartigue, et al 2007.

building of genes and in the magnification of genes from samples by PCR, there is no scientific reason to put a lower limit on the size of the CSNA to include in the system.

We next wished to determine the smallest practical length of CSNA needed to synthesize a gene or genome from oligos. To do so, we examined the open literature on the use of synthetic oligos to construct a virus de novo. We found two reports where the entire genome of a virus had been created from CSNA.[98,99] In both cases, the CSNA was constructed to match the genomic sequence of a known virus. One publication reconstructed the polio virus, which is approximately 7kb long, the other reconstructed the phi X174 bacteriophage (a virus that infects only bacteria), which is approximately 6kbp long. These accomplishments were published in 2002 and 2003, respectively, and we found nothing more recent where a whole genome has been constructed. Although, in 2005 Drew Endy's group at MIT replaced the first 12,000 nucleotides (out of the nearly 40,000bp genome) of bacteriophage T7 with synthetic sequence of their own design.[100] Their goal was to remake part of the virus so that they could better understand how the virus functioned rather than to make an agent de novo.

Today, companies exist with the capability to synthesize DNA oligos with a specified sequence of DNA up to approximately 150 nucleotides long, or, by joining several smaller pieces, genes up to 35,000bp.[101] Given that long oligos are made by piecing together shorter, overlapping strands, we set out to determine how short a piece of DNA can be and still enable the construction of longer (gene or genome length) pieces of DNA. Close examination of the articles mentioned above demonstrate that a viral genome can be built using oligos only 42nt long. Therefore, the purchase of custom DNA 42-mers obviously has to be considered enabling for the synthesis of viruses.

To develop an estimate for the minimum overlap length required to build a longer pieces of DNA, we used primer design tools found at Promega.com to determine three different estimates for the melting temperature of each primer (considering just the portion of the oligos that overlap). The annealing temperature is generally considered to be approximately 5°C lower than the melting temperature. We found that primers consisting of 15 nucleotides or less tend to require annealing temperatures much lower than recommended, making base mismatches much more likely, resulting in an incorrect sequence for the duplex formed by the annealing reaction. Because it is not likely that the thermodynamic profile would allow oligos with less than 15bp overlaps to successfully anneal on a frequent basis to form an entire gene, it is unlikely that anyone would be able to construct a virus or a gene from oligos less than 30nt long. Furthermore, one of us, Dr. Jennifer Byers, who has hands-on experience in synthetic gene synthesis for molecular work in *Entamoeba invadens*, has observed that overhang length greater than 15 bases was required to get any product when building a gene, in practice.[102] Our analysis is reinforced by data from John Mulligan from Blue Heron, who indicates that the industry norm is to use oligos 35-60 bases long for gene synthesis, which suggests that oligos 35nt or shorter would not be likely used for constructing a larger section of DNA.[103]

---

[98] Cello, J, et al, "Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template" *Science*, **297** (1016), 2002.

[99] Smith, HO, et al, "Generating a synthetic genome by whole genome assembly: phi X174 bacteriophage from synthetic oligonucleotides" *PNAS* **100** (26), 2003.

[100] Chan L, et al. "Refactoring Bacteriophage T7" *Molecular Systems Biology*, doi:10.1038/msb4100025, 2005.

[101] Codon Devices fabricated a 35,000nt gene in July of 2006 as reported by Herper, M, "Biggest DNA Ever Made" *Forbes*, July 13, 2006.

[102] Byers, J unpublished data.

[103] Personal communication with John Mulligan, CEO of Blue Heron 6/15/07.

The Cost and Benefit of Excluding Oligos under a Threshold Size from a Screening System

Through the analysis above, we are able to understand what length oligos enable PCR and gene/genome synthesis of activities. To understand the cost of including oligos of various lengths in a screening system, we needed to understand how length of oligo related to order volume. [Commercial proprietary information redacted.] It should be noted that the growing popularity of direct purchase of genes for cloning (instead of PCR amplification of a gene) and of DNA-directed RNA interference is likely to increase the average length of CSNA ordered in the future.

[Commercial proprietary information redacted.] These data about oligo sizes correspond well with the data we have collected from interviews of oligo synthesis companies. By volume, companies we interviewed claim that the average size of oligos sold was between 25-30nt.

[Commercial proprietary information redacted in figure.]

Figure 1. Redacted

These data indicate that the vast majority of oligos are too short to be used directly in gene (or virus) synthesis processes. If PCR-dependent acquisition of genes is not a focus on the system, the vast majority of the oligos ordered today would not need to be screened, reducing operating cost and frequency of false-hits in primary screening. We further divided order volume by DNA length (Figure 2, below) and we found that, indeed very few oligos are too short to be used in PCR. [Commercial proprietary information redacted.] This reinforces the fact that even if we could identify a concrete practical lower limit of oligo size for PCR, excluding oligos shorter than this limit would provide very little benefit in terms of reducing cost or false positives.

[Commercial proprietary information redacted.]

Figure 2. Redacted

[Commercial proprietary information redacted.] Therefore, if a system were concerned only with gene (or virus) synthesis and not PCR, the majority of customers today would fall outside of the system.

[Commercial proprietary information redacted.]

Figure 3. Redacted

[Commercial proprietary information redacted.]

[Commercial proprietary information redacted.]

Figure 4. Redacted

Summary Cost and Benefits of Screening Related to Oligo Length

By determining that it is unlikely for oligonucleotides shorter than 35nt long to be used for gene and virus synthesis, we have potentially eliminated a portion of oligos that would need to be screened. Although these oligos may be used to PCR amplify gene sequences associated with pathogens, it is important to note that one must already poses a sample of the pathogen before amplification can take place. The effect of not screening oligos shorter than 35nt is that there will likely be fewer false hits in a notional screening program, a fact that would be welcomed by all companies in the industry. We believe this to be so for two

reasons. First, the companies would screen fewer sequences, providing fewer chances for false hits to occur. [Commercial proprietary information redacted.] Second, in a system that excludes short oligos, the sequences screened would be longer, and therefore more likely to be unique, thereby reducing the prevalence of false hits even further. Moreover, fewer scientists would be burdened by a system that only screens oligos greater than 35nt as the majority of researchers today order only oligos shorter than that limit. [Commercial proprietary information redacted.] A system focusing only on the longer pieces of CSNA would therefore burden much less of the community than a system that included CSNA of any length.

The downside of not screening oligos smaller than 35nt is that it has little hope of identifying and preventing the malicious use of PCR for the purpose of amplifying genes associated with pathogenicity from samples for the purpose of enhancing the pathogenicity of commonplace microbes.

DNA Modifications and Analogs

The CSNA synthesis industry manufactures more than just DNA as it is found naturally in the body. For various molecular biology applications, researchers desire specific modifications to the oligos they order. Also, commonly the most convenient reagent for an experiment with DNA is not DNA itself, but a nucleic acid analog. In this section, we briefly overview the most common of these modifications and consider if these oligos need to be included in our system. As noted in the Scope section of the Introduction, we were not tasked to perform a full threat analysis, so this section is a brief and simple exploration of the topic.

DNA Modifications

Oligos are often modified at various sites with small molecules to imbue a desired property. A brief accounting of the common modifications is below:

- phosphorylation
- addition of an acrydite group
- addition of a digoxigenin group
- addition of a cholesterol molecule
- addition of a free amino group
- addition of a biotin
- addition of a thiol group
- addition of a fluorophore or quencher
- addition of spacers either at the ends of the oligo or in the place of a base
- addition of a inverted deoxy-thymine cap
- substitution of a 2-aminopurine for an adenine
- substitution of a 2,6-diaminopurine for an adenine
- substitution of a bromo-deoxyuridine for a thymine
- substitution of a deoxy-inosine or 5-nitro-indol for any base
- substitution of a iso-deoxy-cytosine or iso-deoxy-guanine for cytosine or gaunine
- termination by a di-deoxy base
- branched DNA

Some of these modifications would severely limit the ability of the oligos to be used in gene cloning via PCR or in gene synthesis via annealing and ligation of overhangs. For instance, the presence of a phosphate or the absence of a free oxygen (as in the di-deoxy bases) on the 3’ end of an oligo will prevent the oligo’s extension by DNA polymerase, which requires a free oxygen to build the growing DNA chain. Many additions may physically hinder the activity of either DNA polymerase or DNA ligase. However,

many modifications, especially those placed at the 5’ end of the oligo may not hinder the ability to perform PCR or any other critical operations at all.

The question arises then: can these modified oligos be excluded from the system? None of these modifications is “desirable” for the use of oligos in gene cloning or fabrication (except for 5’ phosphorylation and certain base substitutions). However, some of these modifications can be easily removed by the action of a specific enzyme without altering the rest of the oligo. In contrast, many modifications cannot be removed without chemically altering the parent oligo. That being said, these modifications can still be removed from the oligo if the customer intentionally orders an oligo that is longer than desired and uses a restriction endonuclease to remove the part of the oligo with the modification. Clearly, it is easier for an adversary not to order modified oligos in the first place and avoid these troublesome steps. However, at this stage, we cannot say with certainty that any type of modification renders an oligo completely unusable for PCR or gene fabrication. Furthermore, some modifications, like the substitution of universal bases, may be desirable when attempting to amplify a gene from an environmental isolate.

DNA Analogs

For various applications, usually to boost the strength of binding of short oligos to their target or to improve stability of the oligo in vivo, some DNA-like molecules have been developed. Several analogs replace the sugar-phosphate backbone with other linkages, for instance, a phophorthioate bond, which substitutes sulfur for the phosphorus in the bond and thereby inhibits degradation but not elongation. Other analogs include locked nucleic acids, peptide-nucleic acids and morpholinos (a DNA-like molecule that uses morpholine rings instead of a carbohydrate backbone).

Some of these molecules, such as peptide-nucleic acids and morpholinos, cannot be extended by DNA polymerase and therefore cannot be used for PCR. Further, some cannot be recognized by DNA ligase, so cannot be used for gene fabrication. Also, some analogs are poorly soluble in water and therefore problems with aggregation will prevent the use of long oligos of these molecules for gene synthesis. Perhaps most importantly, DNA analogs are likely not to be accepted by the normal cellular machinery and so must be removed before introduction into a cell. Bear in mind, however, that these molecules are typically not meant for incorporation into genetic material in vivo (and are in fact used to interfere with normal DNA or RNA processing), so their behavior as part of a gene has not, to our knowledge, been studied.

When a researcher wants to incorporate these DNA analogs into a reagent and perform ligation or extension reactions, a DNA-analog hybrid reagent is ordered, which contains some bases of DNA and some of the analog. As long as normal DNA is on the 3’ and/or 5’ end of the molecule, extension and ligation can occur normally. Furthermore, the analog component of a long hybrid oligo could be removed by restriction endonucleases as described above. Alternatively, the hybrid molecules could be used as primers for PCR and then the hybrid component removed by endonucleases after amplification.

RNA, chemically similar to DNA except that RNA has a free oxygen on the 2’ carbon of the ribose, is a DNA analog that is abundant in the body. Also, one can buy RNA oligos of any desired sequences much as one can for DNA oligos, typically up to 50nt long. For some applications of concern, RNA oligos may be ideally suited. Clearly, for the de novo synthesis of an RNA virus (that is, a virus that uses RNA as its genetic material instead of DNA) an adversary may try to use RNA oligos directly. However, we estimate that, even for some RNA viruses, it may be advantageous to make a copy of the viral genome in DNA and transcribe the RNA genome from a more stable and more practical DNA template. In fact, this is the

approach that Wimmer's group used to synthesize the poliovirus, which is an RNA virus.[104] There are several reasons, both molecular and cell biological, for not using RNA oligos in the place of DNA oligos for PCR or gene fabrication when an RNA virus is not the desired product (for instance, the inferior stability of RNA and RNA-DNA dimers, and inability of thermostable polymerases to extend RNA primers, etc). However, with some modifications, RNA oligos can be used to direct the synthesis of homologous DNA oligos using a viral enzyme, reverse transcriptase. In fact, this procedure is commonly used to take a cellular pool of RNA molecules and obtain a more stable and more useful pool of DNA molecules for genomics research. Therefore, although RNA oligos may be less useful directly for PCR or gene fabrication, it can be used to obtain DNA oligos relatively easily.

[104] Cello, et al, 2002.

Degenerate Oligos

A degenerate oligo set is a pool of oligos in which a random base will be substituted for bases throughout the primer. That is, any base has a chance of being randomly substituted for the other bases and a set contains many similar oligos that differ from each other by two bases on average. These oligos are used for specialized applications, for instance in PCR when amplifying a gene from an environmental isolate where specific allelic information is unavailable. The variety in a degenerate primer set enables the amplification of genes that differs modestly from a "canonical" gene from that species due to genetic drift. Therefore, these primers are particularly useful to a researcher attempting to amplify a particular gene from an environmental sample. We note that modified oligos that substitute a universal base, such as 5-nitro-indol as above in the section describing DNA modifications, for a natural base, have many of the same advantages as degenerate primers.

Degenerate oligos are not well-suited for the fabrication of genes because the random bases incorporated throughout their length will greatly complicate the annealing steps, thereby lowering their melting temperature. Even if these oligos could be used in the fabrication of a gene, the random substitutions would incorporate many mutations in the final genes, reducing the likelihood that the product is useful.

The Benefit of Inclusion of Non-standard Oligos in a Screening System

Many oligo modifications and DNA analogs are unsuited for direct use in PCR and gene (and therefore, genome) fabrication. However, many of these modified oligos, RNA oligos and some DNA-analog hybrid molecules could be modified via enzymatic action or restriction endonucleases to result in unmodified DNA oligos perfectly suitable for both PCR and gene fabrication. We have not conducted a full process assessment at this stage to determine the exact comparative difficulty using these unusual oligos to arrive at desired molecular products, or indeed to identify exactly which modifications or analogs would pose no issue to those with malicious intent. Further study would enable us to determine the relative reduction of the risk represented by each modification or analog to identify the exact benefit of including a particular type of oligo in the system. Clearly, if a decision is made to exclude non-standard oligos from a screening system, a thorough risk assessment must be performed that considers how easily the non-standard oligo could be used to undertake activities of concern.

Our present analysis enables us to conclude that the use of these modified oligos and DNA analogs would add at least one step, and perhaps several complex steps, in the amplification of a gene or the fabrication of a genome beyond that required when using unmodified DNA oligos. RNA, although unsuitable for direct use in gene fabrication or PCR, could be used to drive the synthesis of the appropriate DNA reagents and may be a suitable replacement for DNA when an RNA virus is the final product. However, we can state that some DNA analogs (specifically, morpholinos and peptide-nucleic acids), when not in a hybrid DNA-analog molecule, cannot be used in PCR or gene synthesis. Other analogs, such as oligos with a phosphorthioate backbone, could be used as primers in PCR.

It must be kept in mind, however, that the only reason to use DNA analogs or modified bases in gene fabrication or PCR (with some exceptions like degenerate primers, oligos with universal bases and perhaps RNA for the synthesis of negative-sense RNA viruses) is to avoid detection by a screening system. Each one of these non-standard oligos has a disadvantage in comparison to unmodified DNA oligos when considering the complexity and likelihood of success of the protocol to arrive at the desired products. Therefore, the inclusion of these molecules in a screening system only makes sense if we believe that the adversary is aware of (and feels confident in the abilities of) CSNA screening systems and tries to subvert detection by using sub-optimal alternatives.

In contrast, a researcher could order degenerate oligos or oligos in which universal bases are substituted for natural bases, as a first choice when attempting to amplify and isolate a desired gene from an environmental (or medical) sample. However, due to complexities in getting partially random oligos to anneal predictably, these molecules are not suited for gene fabrication protocols.

The Cost of Inclusion of Non-standard Oligos

Given that, except for degenerate oligos, an adversary would only use these non-standard oligos if they were aware of and wanted to avoid a screening system, we must then calculate the cost of including these molecules in a screening system. [Commercial proprietary information redacted.]

[Commercial proprietary information redacted.]

Figure 5. Redacted

Given data collected in our interviews with CSNA providers, more than 90 percent of the CSNA produced (by volume) is for unmodified DNA oligos (though it is often purified). Therefore, when considering only the total volume of unmodified DNA oligos compared to non-standard oligos, including the screening of these molecules will only marginally affect the quantity of oligos that need to be screened and therefore would have minimal effect upon the cost and false hit frequency. However, as described in the Supplier Review, below, many companies specialize in the production of non-standard oligos and make very little, if any, unmodified DNA oligos. Companies that make unmodified DNA oligos comprise the majority of the CSNA market and most of their revenue comes from producing unmodified DNA oligos. Other CSNA companies have left the production of unmodified DNA oligos to these market leaders and focus on "niche" products. In fact, we identified at least five companies through interviews that do not make unmodified DNA oligos or genes at all. Given that all of the companies we could not interview were small, and together comprise less than 20 percent of the market, we suspect that many, if not most, of the smaller companies in fact make very little, if any unmodified DNA oligos or genes. Therefore, the exclusion of these non-standard oligos from the system would free many companies from the costs of participation in the system. These companies are primarily small and would therefore be less able to bear additional screening costs.

Non-standard Formats

Several providers of CSNA cater to specialized needs of their customers, and so only provide CSNA in a specific format that may be unsuitable for use in activities of concern. For example, one company we interviewed only produces oligonucleotides in very large batches (at least one milligram per oligo ordered, compared to typical order quantities of a picogram—10^{-12}g—per oligo ordered). Because the quantities this company delivers are large, they charge at least $1,000 per oligo, compared to less than $50 for similar, purified product from other suppliers. If a hostile actor only needed two oligos for primers in order to amplify a gene via PCR, this may be a feasible source for those primers as the increase in expense is not that great. But, if the weaponeer needs a series of 300 oligos to synthesize a gene or a viral genome de novo, this approach would be extremely costly compared to ordering oligos in standard quantities. Although we note that $300,000 is well within the reach of some groups who wish to inflict mass casualties, this expense is only rational if the weaponeer has confidence that a CSNA screening system would catch the activity if he ordered oligos in standard formats.

In addition to quantity, DNA is sometimes delivered, not as a powder in a tube or a 96-well plate, but attached to a substrate. For example, CSNA may be delivered attached to a micro-array chip, where many (up to a few hundred thousand) oligos may be attached to one surface. These oligos are variable in length, and are anchored to the array substrate in a manner that makes release of the oligo from the array

difficult. In fact, current technology does not allow for oligos to be easily removed from array substrates. Although several companies are currently pursuing how to utilize array-based oligos for small-scale synthesis, currently progress has been stymied by the difficulty of removing the oligos from the array. Low tech methods that would be available to a hostile actor for removing oligos from an array include scraping the array or some other "brute force" means, or designing the array to include restriction endonuclease recognition sites near the base of each oligo of interest on the array. Scraping or other brute force method would likely damage or contaminate the oligos to a point where they were not be useful for gene synthesis or PCR. If some enzymatic means were used, such as restriction endonucleases, the array could be designed to release desired oligos in a stepwise fashion after each restriction enzyme is used (for instance, the first desired oligo would be designed with an EcoR1 restriction site, the enzyme used and and then harvested from the array, then the second desired oligo could be designed with a BamH1 site and harvested, etc). This scenario shows that an extremely determined individual could come up with creative methods to possibly surmount some technological barriers, but the amount of work that is required is dramatically harder than other methods of obtaining oligos (like, in this case, purchasing their own DNA synthesizer) and therefore a more likely option. Furthermore, this convoluted method would only be attempted by actors who are aware of and have confidence in the ability of the system to prevent the acquisition of their reagents if oligos were ordered in a standard format.

DNA can also be ordered attached to small beads of polystyrene or silica. If the oligos are attached to the bead by an uncleavable linker, the difficulties of removing oligos from the beads for use in PCR or gene or virus synthesis are similar to oligos attached to a micro-array. However, some oligos are attached to beads with cleavable linkers, designed to enable the easy recovery of oligos from the beads. In this case, the harvesting of oligos is somewhat simple. However, once again, an adversary would only order oligos attached to beads if he knew about and wished to confound a possible screening system.

Conclusions

A CSNA screening system that focuses on only the fabrication of genes and genomes (compared to the use of PCR to amplify genes from an environmental sample) need only include CSNA products greater than 35nt. The exclusion of shorter nucleotides, while vitiating the ability of the system to prevent the misuse of PCR, would drastically reduce the operating cost of the system by screening only a minority of the total CSNA products made each year, inconveniencing only a minority of molecular biologists, and simultaneously generating fewer false hits per order screened.

Modified oligos, CSNA in non-standard formats, and DNA analogs (with few exceptions) are not well-suited for gene cloning or genome fabrication and therefore an adversary would only use them for these purposes if they were trying to evade a screening system. Degenerate oligos are unsuited for gene fabrication but may be very desirable when cloning a gene from an environmental or medical isolate. Although these non-standard oligos comprise a small fraction of the oligo market, many companies only produce non-standard oligos and focusing the efforts of a screening system on fewer companies may enable the most precise application of resources where they are most needed.

The Suppliers

Synthetic gene and oligo providers were contacted in an attempt to understand each industry's products and procedures. Company representatives were questioned about sequence and customer screening best practices, and opinions regarding oversight were sought. Additionally, information was gathered to determine how a screening system could be implemented and the impact it would have on the suppliers'

operations. Using strategically targeted internet searches and personal communications with industry leaders and synthetic biology researchers, Gryphon Scientific identified 29 US companies currently producing oligos and 20 US companies that currently make synthetic genes (note that nine companies are counted in both groups as they produce both products). Importantly, after identifying the companies, we called them to validate that they were still in business and still offered custom oligos. Of the companies initially identified through internet search or by being named as competitors by people in the field, approximately 10 have either gone out of business or no longer produce CSNA, underscoring the consolidation occurring in the industry. Interviews were undertaken from companies that still make CSNA to gather information about the company's sales, as well as their practices and opinions regarding screening. Results and opinions from each industry will be reported separately in this document as opinions and current practices vary dramatically between the two industries.

Synthetic gene and oligo provider participation in our study can be seen in Figure 6. Of the companies we identified, 40 percent(primarily small or niche companies) were non-responsive (despite several attempts to contact them by phone and email) or declined to speak with us for a variety of reasons including time constraints, confidentiality concerns or simple unwillingness. We believe that reluctance to even discuss this topic with us demonstrates a marked hostility towards potential government involvement in their industry. However, we were able to obtain data and opinions from 60 percent of the companies we identified, including the industry leaders such as Integrated DNA Technologies, Invitrogen, Sigma-Genosys, Operon, Blue Heron, Codon Devices and DNA 2.0.[105] Based on information provided by these major suppliers and the others that were interviewed, we found that these major players dominate the market. Because we interviewed all major players, we therefore estimate that our survey represents greater than 85 percent of the value and volume of the US-based oligo and synthetic gene supply. Given the dominance of US CSNA producers in both the gene and oligo market, and the size of the US-based pharmaceutical and biotechnology industries compared to foreign industry, we estimate that these companies represent significantly more than half of the world CSNA supply as well.

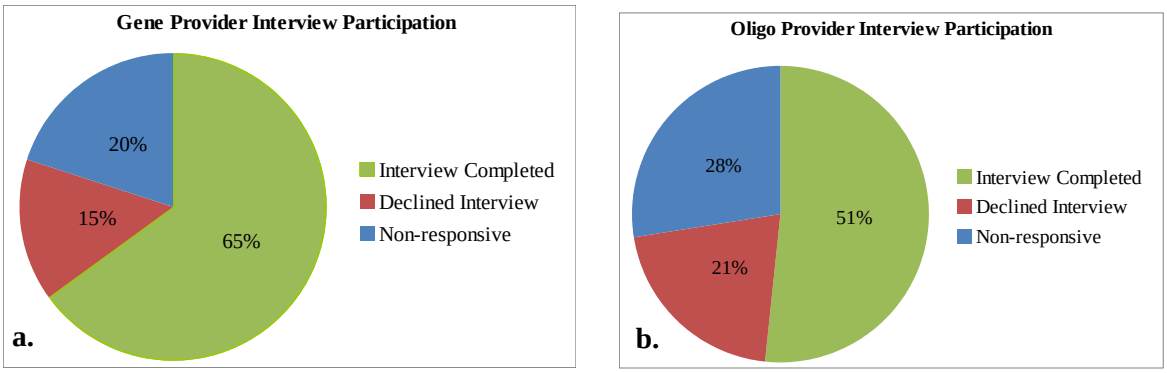


Figure 6. Breakdown of interview participation by the gene (a) and oligo (b) industries.

The awareness of the need for oversight of the industry varied among interviewees. Many companies we spoke with in the synthetic gene industry expressed an awareness of the potential for misuse of synthetic DNA technology. Several representatives from the synthetic gene industry spoke about their involvement with the International Consortium for Polynucleotide Synthesis (ICPS) and the efforts this organization has put forth to promote self-regulation of the gene synthesis industry. Conversely, many in the oligo synthesis industry feel the potential for misuse of their product is small and regulation therefore unnecessary.

[105] For a complete list of companies please see Appendix I.

This industry overview is intended to provide a picture of each industry's typical product, their current practices regarding order processing and screening, as well as each industry's view about screening and the future of synthetic biology. The information we collected was gathered from our interviews, unless otherwise noted.

Typical Orders

It is not surprising that the gene synthesis industry and the oligo industry have differing views about the potential for misuse of their products. Although both industries sell synthesized nucleic acids, the size, quantity of products sold, and time required to produce said product is very different. Therefore, we discuss each component of the CSNA industry separately in the section below.

Synthetic Genes

By their nature, genes are larger than oligos and do not contain chemical modifications sometimes added to oligos. Synthetic genes can range in size from 150bp (for a mini gene) to over one hundred kbp for large genes. During the course of our industry interviews, 52kb was the largest single product anyone mentioned producing, but several companies felt that they could produce longer genes. In their opinion, the current limiting factor was demand, not technology. Those we spoke with in the industry estimated the average synthetic gene purchase is on the order of one to two kilobase-pairs. Genes of this size are often constructed in one to two weeks, while larger genes or those that are complex can take several months to complete. While many of the larger gene companies we spoke with make a few hundred or more genes per week, smaller players in the gene synthesis industry may make as few as 10-20 genes per week.

Oligos

Unlike synthetic genes which are often thousands of bases long, oligos are typically much smaller, ranging in size from two to 150 bases.[106] With the average oligo order containing less than 30 nucleotides, orders can often be received, synthesized and shipped all in the same day. [Commercial proprietary information redacted.] In contrast to synthetic gene providers, most oligo companies make thousands, tens of thousands or even a hundred thousand or more oligos per week. Of these orders, our industry contacts report that approximately 90 percent are for "vanilla" oligos, that is, un-modified, small-scale orders. For some applications, oligos, even those of the vanilla variety, require a purification procedure such as HPLC or PAGE. The other 10 percent of orders include those for oligos that are biotinylated, phosphorylated or otherwise modified with a tag, those that are ordered in large (milligram) quantities, and those for non-DNA oligos (including RNA oligos and morpholinos). Using [Commercial proprietary information redacted] we can estimate that approximately 25 percent of unmodified oligos are purified.

Market Value

By combining information from our industry interviews with [Commercial proprietary information redacted.], we estimated the value of the synthetic DNA industry as follows: Industry leaders report that approximately 26 million oligos are synthesized each year with an average oligo length between 25 and 30 nucleotides. Using an average price of $0.35 per nt (our calculated average for 2007) coupled with the estimate that 25 percent of oligos require purification (at a cost of approximately $50/oligo) and that 10 percent of oligos receive a modification (estimated to be around $150/oligo), we calculate the oligo

[106] Rarely, oligos larger than 120 bases are produced, however, due to the inefficiencies in the synthesis process, oligos greater than 120 bases are difficult to fabricate without piecing together smaller oligos. Also, it is rare for customers to order oligos shorter than 10 bases.

market value to be in the neighborhood of $950 million globally. To determine the value of the entire synthetic DNA industry the gene synthesis industry was also taken into account. The number of genes synthesized each year (~50,000) is relatively small compared to the oligos synthesized. With the average 1kbp gene costing around $1,000 (for a total gene market estimated at $50 million), we estimate that the entire CSNA industry is valued at approximately $1 billion/year globally. [Commercial proprietary information redacted.]

Order Processing Procedures

Below we detail the order processing procedures used in the synthetic gene and oligo industries. From receiving orders, to synthesizing products, to shipping goods, our industry contacts walked us through their procedures to help us understand how sequence screening is, or might be, incorporated into their current practices.

Synthetic Genes

A great deal of variation exists within the synthetic gene industry with regard to order processing. While a few of the larger companies indicated that some orders are received via a click-through internet contract, others pointed out that ordering synthetic genes can be more complicated. Due to the complexity and expense sometimes associated with gene synthesis, orders may involve negotiations, phone conversations, and/or email correspondence before the product is delivered. Customers may be asked questions such as, what protein they are trying to express, what organism the gene will be expressed in, or the exact sequence being requested. Several synthetic gene companies we spoke with require customers to fill out a disclosure of hazards and biosafety concerns before accepting an order. Once orders are completed, design and synthesis can begin.

Many in the synthetic gene industry use internal software to evaluate sequences prior to synthesis (and in some cases prior to accepting an order). The software, which is primarily used to aid in gene design (searching for hairpin structures, high GC content and repeat regions), is also often employed to screen sequences for elements associated with pathogenicity. Those companies that screen for such sequences say the screen requires almost no additional resources unless a pathogenic element is identified, in which case follow-up conversations are required. Once gene design is complete, typically in less than a day, synthesis begins. While most gene providers synthesize oligos (which are used to build the gene) themselves, a few purchase oligos synthesized by another company with which they build their genes. A simple gene can then be constructed in as little as 1 -2 weeks with large, complex genes often taking up to two months or longer for completion. Completed genes may be shipped as naked, purified DNA or in a bacterial sample transformed with a gene-containing plasmid. Shipments are typically sent out once a day using an overnight carrier such as Fed Ex, UPS or DHL.

Oligos

Unlike synthetic gene orders, oligo orders can be placed relatively simply. Most orders are submitted by entering the desired sequence on an internet site or in an email with a small minority of orders received via fax or phone. While the minority of orders that are not received electronically must be entered by hand, the majority of orders, which are received electronically, typically enter a queue for automated processing and may begin to be synthesized within 15-60 minutes of their receipt. Although no companies we spoke with currently screen oligo sequences for pathogen-containing elements, this 15-60 minute period may provide a window where screening could be implemented prior to synthesis. Oligos are desalted, lyophilized and may be shipped in small tubes or, when a large number of oligos are ordered, 96-well plates (or other formats). Unlike synthetic gene companies which do not offer same-day shipping, nearly every oligo company we spoke with offers same-day shipping on orders placed before a certain time of day (often 3:00 or 4:00 pm). Although a few companies reported multiple shipping

pickups per day, most shipped just once a day (typically in the early evening) using FedEx, UPS, DHL or another overnight carrier.

Client Base

Synthetic gene construction and oligo synthesis are international businesses. Companies are based in many countries (including the expected players, the United States and Europe, to the unexpected, India, Korea, China, and Iran) and their clients are found all over the globe. Many of the US-based companies boasted synthesis facilities in multiple countries. US Companies we spoke with reported selling CSNA to customers in countries all over the globe including, Israel, Egypt, Japan, China, India, Australia, Brazil, Ecuador and Venezuela, with *most* prohibiting sales only to those in Sudan, Syria, Iran, North Korea, and Cuba. We found it interesting that companies are shipping oligos that may or may not be associated with pathogenic components to companies that should not receive dual-use biotechnology items (because they are outside of the Australia Group, such as Israel and Egypt).

Synthetic Genes

Companies we interviewed all had at least one US branch and most were headquartered in this country (we interviewed some foreign companies only to gather best practices related to screening). Based on this, it is not surprising that the synthetic gene suppliers we spoke with indicated that between 70 percent and 90 percent of their customers are US based, with the majority of other clients being based in Europe or industrialized Asia. Pharmaceutical and biotechnology companies make up the majority (70-80 percent) of synthetic gene customers; government-run research organizations (like the NIH and national laboratories) are this industry's next biggest client. As this industry matures and more scientists become aware of potential time and cost savings attainable by integrating synthetic genes into their experiments, it is likely that academe will become a significant portion of this industry's customer base.

Oligos

The picture observed in the oligo industry for international versus domestic customers is similar to that observed in the gene synthesis industry with 70-90 percent of companies based in the United States. However, unlike the gene synthesis industry only about 40-50 percent of oligo synthesis business comes from pharmaceutical and biotechnology companies. One factor that might explain this discrepancy is price. At approximately $0.35/base an average (30nt) oligo costs just over $10.00, these products are affordable for nearly anyone in the life sciences. On the other hand, genes are still relatively expensive to purchase, averaging around $1,000 for a 1 kb gene, which may be prohibitively expensive for some in academia. Additionally, oligos can be synthesized cheaply and with relative ease by those possessing a DNA synthesizer.

Sequence Screening Practices and Opinions

One of the biggest differences between the oligo synthesis industry and the gene synthesis industry is the view each takes on screening. Of the 13 US-based companies and six international companies we spoke with that make synthetic genes only three (all US-based) do NOT currently screen sequences for pathogenic elements. This fact should not be interpreted to mean that foreign gene synthesis companies are more engaged in biosecurity; it is likely that only the most proactive foreign companies chose to speak with us, biasing our survey of foreign companies towards those that screen. In contrast, no company we identified screens oligos for content (though many do screen customer information).

Synthetic Genes

One practice used by several gene companies is the requirement that all customers declare whether their sequences may be hazardous. Many synthesis providers use a BLAST-based algorithm to screen all sequences against an internal database that recognizes pathogens or pathogenic elements. The exact sequences each company screens against differs between companies, although many companies we spoke with indicated they screen for sequences from organisms on the Commerce Control, Select Agent, or CDC Bioterrorism Agents/Diseases lists.

A few gene companies have a business model that ensures they know exactly what they are making. Rather than constructing a gene based on sequence, the company designs and synthesizes genes based on the protein the customer needs produced. Because the company knows what protein is being made and designs the gene, the possibility of making something hazardous without their knowledge is reduced. Of the US companies that screen sequences, two take advantage of the Craic Computing program BlackWatch to screen for pathogens while the other eight use programs they designed themselves (most of which are BLAST based). Of the six international companies interviewed, two use BlackWatch and the other four use a BLAST based program. Figure 7 depicts screening methods used by US and foreign based gene synthesis companies.

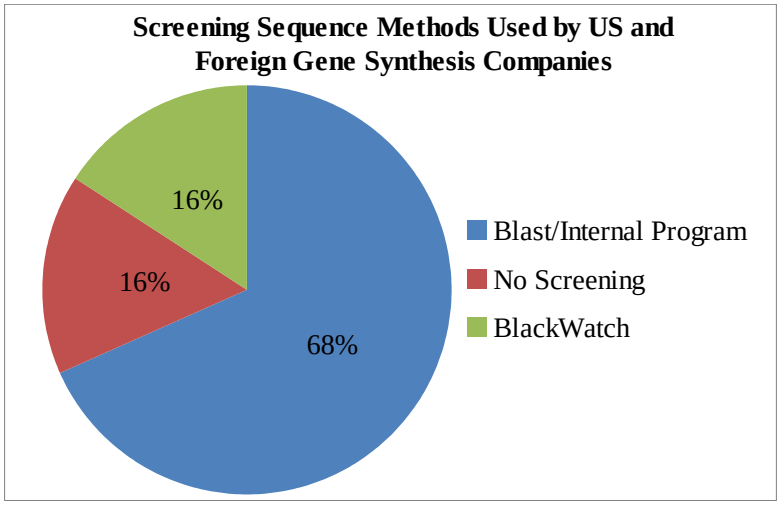


Figure 7. Sequence screening methods used by US-based and foreign gene synthesis companies.

Most gene synthesizing companies in both the United States and abroad had a positive or neutral attitude toward screening (see Figure 8a). One US gene synthesis company had this to say: "The Government should have oversight of potential sequences derived from pathogenic origin [sic] that are synthesized by companies. If such sequences are synthesized…the system should be so designed that this information is stored in the Federal database. This information may never need to be used, but it is better to be cautious."[107]

Although most gene companies we spoke with screen sequences, some, even those that screen do not feel that it is the most effective way to prevent misuse of the technology. Some note that it is the customer, and not the sequence we should be concerned about. For example, many customers, such as those working on vaccine design or pathogen diagnostics have a legitimate use for synthetic DNA containing pathogenic elements. Screening sequences requested by these companies will certainly detect orders for sequences from pathogens; however, these customers should not be prohibited from purchasing said

[107] All interviewees spoke to us understanding that their specific opinions or practices would not be associated with their name or company specifically.

sequences. Other concerns brought up by representatives of the synthetic gene industry included concerns that there is no perfect screening tool (many mentioned that BLAST algorithms may yield false positives or false negatives).

The most prominent concern about sequence screening brought up by those in the gene synthesis industry was how the screening would be regulated. Nearly everyone we spoke with opposed sending sequences outside of their firewall for third party (such as government) screening. Many felt customers (especially pharmaceutical and biotechnology customers) would choose to make their own genes rather than chance the release of proprietary information when their sequences are released to a third party. Despite the concerns addressed above, most representatives we spoke with from the gene synthesis industry (83 percent) are not opposed to sequence screening provided it is handled internally unless a problem arises.

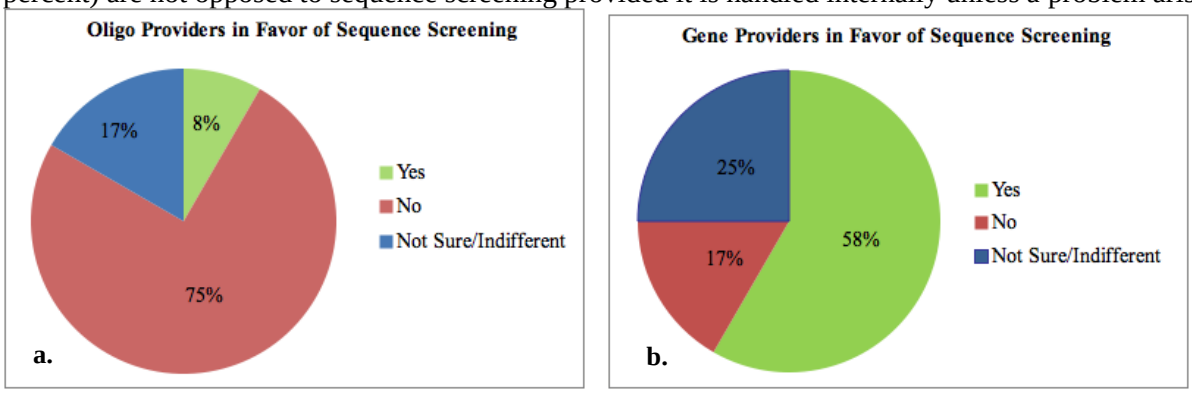


Figure 8. Opinions of oligo (a) and gene (b) providers regarding screening sequences for pathogenic elements.

Oligos

Although most gene synthesis companies are already screening for pathogens or are open to the idea of doing so, most who synthesize oligos (75 percent) are against screening their products (Figure 8b). "WE DON'T SCREEN. IT'S NONE OF OUR BUSINESS" (emphasis theirs in a text response) reports one oligo synthesis company, adding "we put the sequence on the machine, make it and ship it. I know some are pathogen sequences because some of our customers are diagnostic companies that are developing diagnostics for pathogens. But, other than that we don't care and do maintain our customer's confidentiality at all levels!"[108]

Others in the industry are more open to the idea, but still feel that screening of sequences is excessive or impractical. "The main applications for our product are for PCR and sequencing. With both eukaryotic and prokaryotic organisms there is a tremendous amount of similar genetic material within any given 30bp region. It would simply not be effective to try to screen sequences for a match to a specific pathogenic species" states one oligo provider. Some note that trying to regulate the oligo industry to prevent hostile actors from synthesizing dangerous organisms would be on par with regulating the steel industry to prevent the illegal manufacture of guns. Additionally, several small oligo companies feared that reporting every sequence that matched a pathogen would be debilitating because of the number of resources and computing power it would require. Citing this concern, one oligo industry representative expressed concern that attempts to regulate the oligo industry might force companies to move their businesses out of the United States. We note that relocating a business outside of the US would impose a significant disadvantage to oligo producers, beyond the actual cost of moving a facility with large pieces

[108] All interviewees spoke to us understanding that their specific opinions or practices would not be associated with their name or company specifically.

of capital equipment and finding qualified workers, because proximity to customers enables them to meet aggressive delivery schedules. Furthermore, many of the larger oligo providers noted that if required to do so, they could incorporate screening into their system for a minimal cost. As with synthetic gene providers, oligo providers were strongly opposed to third-party (government) screening. Some felt third-party screening would significantly slow down production while others cited customer confidentiality concerns similar to those expressed by the synthetic gene providers.

Of the 15 US companies we interviewed that sell oligos, none currently screen their sequences for pathogenic elements. Interestingly, several companies that make both synthetic genes and oligos choose to screen gene sequences but not oligo sequences for pathogenic elements.

Several companies mentioned that screening oligo sequences would be very difficult because of the large number of oligos that might contain a pathogenic element but have a legitimate use. One oligo provider has this to say: “We make a lot of [oligos] that are designed to DETECT biothreat agents but have absolutely nothing to do with synthesizing genes or assembling segments of an organism. If you compare [sequence] signatures to a database a significant percentage of oligos may "hit". Our customers have included/and include USAMRIID, DOE Labs, NIAID, DHHS, CDC, USDA, FBI, US Navy, US Air Force, US Army, Plum Island and a number of other less public agencies which prefer to keep their sequences proprietary. All of these customers had [oligos] made for detection, none of which expressed an interest in gene synthesis.” For reasons such as this, many in the oligo industry feel the best way to prevent misuse of synthetic DNA is by screening customers (rather than sequences). However, the overwhelming consensus among those in the oligo synthesis industry is that oligos do not pose a significant biosafety risk and do not need to be screened.

Many in the oligo industry argue that screening is not necessary because they feel it would be significantly more difficult to construct a hazardous organism than it would be to obtain a pathogen from nature or a laboratory stock. It is probably with this type of comment in mind that several who proposed oversight of the CSNA industry suggest that such regulation focus on organisms that are particularly difficult to find (such as the 1918 flu virus or Smallpox Virus). Several representatives of the oligo industry noted that the skill required to construct a viable infectious organism using synthetic oligos is available only to sophisticated researchers, a condition which they feel significantly reduces the risk associated with oligos. Furthermore, oligo providers feel that those that wished to cause harm could easily circumvent any regulation within the United States by ordering oligos from another country or by self-synthesis.

Screening Customers

As pointed out by both the synthetic gene and oligo providers, customer screening may help prevent hostile actors from gaining access to synthetic DNA technology. Unlike sequence screening which is strongly opposed by those in the oligo industry, representatives from both the synthetic gene and oligo industries were in favor of or indifferent to customer screening (Figure 9).

Synthetic Genes

Slightly more than 60 percent of the domestic gene synthesis companies we spoke with used some method to screen customers, slightly less than those from the same industry who screen sequences. Those that did screen did so for a variety of reasons using a range of techniques. Many gene synthesis companies we spoke with indicated screening involved “verifying the legitimacy” of their customers. For many “verifying legitimacy” involved nothing more than reviewing the customers shipping and e-mail address, as nearly every company we spoke with indicated that orders from Hotmail or Gmail accounts or those with residential shipping addresses raised a red flag. For other gene synthesis companies, the

practice of verifying legitimacy was more thorough. One gene synthesis company screened customers against the Dun and Bradstreet database. Because not all legitimate customers (such as those associated with a university) appear in the Dun and Bradstreet database, a third party reference is used to screen customers not appearing in the database. Another company that makes both synthetic genes and oligos screens foreign customers using Bridger Insight's "Choice Point" software. This software screens foreign customers against all of the US Commerce Department's "Denied Parties" lists, which contains companies and people suspected of contributing to the proliferation of sensitive technology. Several other companies we spoke with indicated that they only sell genes to companies with tax ID numbers. Although not practical for large gene synthesis companies, one small company we spoke with mentioned that they have only done business with individuals that they know personally or from scientific meetings.
In addition to "verifying the legitimacy" of customers, many gene synthesis companies we spoke with screened customers for their ability to pay. Some companies conducted credit checks while others required payment or partial payment upfront.

As with sequence screening, customer screening was performed across the board in the foreign gene synthesis companies with which we spoke. Most international companies screen primarily against customer shipping and e-mail address. One international company indicated that they screened customers using the manual of German export control (HADDEX) while another cross-checks addresses with university and company directories to ensure their customers are currently employed noting that "for small biotech companies (<100 employee), there could be several rounds of screening in terms of company profile, company research, references, and sequence screening [before an order is confirmed]."

Oligos

Unlike sequence screening which is non-existent among oligo providers, 53 percent of oligo companies we spoke with did some kind of customer screening. Some oligo providers we spoke with attempted to verify the legitimacy of their customers through many of the same means as the gene synthesis providers (e-mail and shipping addresses, Dun and Bradstreet databases, and tax ID numbers). One of the most rigorous customer screening practices was performed by an oligo provider who mentioned that their company takes a detailed view of each of its customers including who they are, what kind of work they are doing and where they are located (no residential addresses are accepted). This oligo company screened all customers against the US Department of Commerce's denied parties lists and the Fedworld site. Although a few of the of the oligo companies we spoke with used customer screen practices aimed at verifying the legitimacy of their customers, most admitted that the customer screening practices were primarily to ensure payment and often involved a credit check.

Like those in the synthetic gene industry, those in the oligo industry expressed a positive or neutral attitude about screening customers (Figure 9b). Most feel that screening customers would require fewer resources and be more effective than screening customers. Additionally customer screening rather than sequence screening would prevent one of the biggest concerns expressed by oligo providers, false positives and false negatives based on sequence data alone and flagging of orders from customers permitted to have sequences from pathogens.

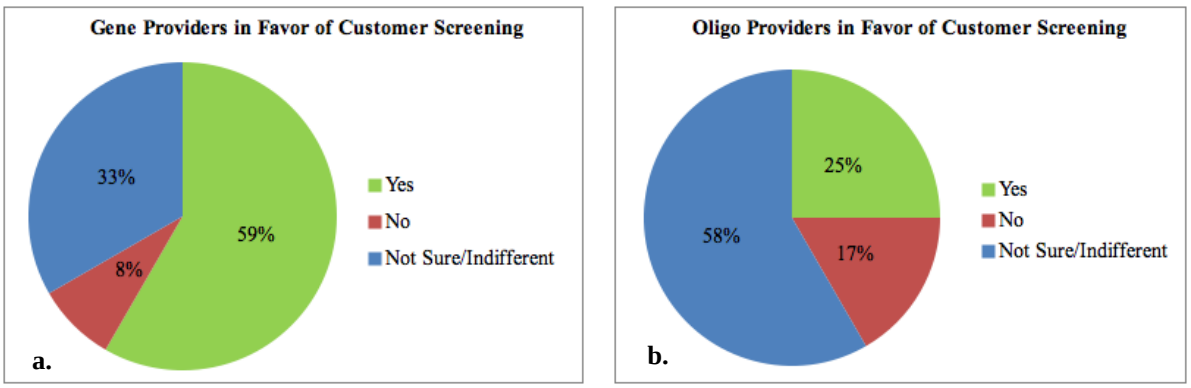


Figure 9. Opinions of gene (a) and oligo (b) providers regarding screening customers.

Post-Screening Intervention

Although many oligo providers screen customers and many gene synthesis companies screen both sequences for pathogenic elements and customers, it seems no one is quite certain about the best protocol to use for handling suspect sequences. This section will address some of the post-screening steps currently employed in the synthetic gene and oligo industries.

Synthetic Genes

Many gene synthesis companies have established their own protocols for handling suspect orders. A few companies we spoke with noted that if an order is suspect, the screener will refer the problem to her supervisor. Once escalated, the company will determine if the gene in question can be synthesized for the customer requesting it, or if further investigation is required. If further investigation is required the company may choose to contact their legal or technical department or to contact the researcher who ordered the gene (and her Principal Investigator) directly, by phone or e-mail, for information about why the gene is being ordered. The company may also choose to contact the biosafety officer at the researcher's institution to obtain additional information. If, after investigation, the company feels the end-user is authorized to receive the product, they will synthesize and ship the gene. Interestingly, one gene synthesis company we spoke with indicated that, because of export regulations, they will not make pathogenic elements for any foreign clients regardless of their credentials.

If the synthesis company determines it is not safe to make the product for that customer, the supplier informs the customer that the order will not be filled. Many synthetic gene companies we spoke with indicated that after such an event they feel they should contact someone at the government about the suspect order, however most companies we spoke with indicated that they have been unable to identify someone to call. Some mentioned that they have contacted the FBI about the suspect orders, but they did not receive feedback that they contacted the appropriate person or that any action was being taken on the order. Although a few interviewees mentioned they had personal contacts in government that they could contact when a sequence was flagged by their internal screen, all said they would like to have an official contact within the government.

Oligos

Despite differences between the oligo and gene synthesis industries, post-customer-screening steps were treated similarly by both industries. Several providers of both synthetic genes and oligos told us that if a customer seems suspicious because of his e-mail or shipping address or because his name appears on a denied party list, further investigative steps will be taken. If the customer can be verified as being

legitimate (for instance the customer only has the same name as someone on the denied parties list), the order will be processed, synthesized and shipped. If the customer's legitimacy cannot be verified, the customer will be informed that his oligo/gene product will not be made. In the event that the customer screening system flags the order, oligo companies indicated they would like to have a government contact to notify.

Figures 10 and 11 present flow charts of how suspect sequences (Figure 10) and suspect customers (Figure 11) are handled at many companies that screen. Note the question marks connecting the "notify authorities" box to boxes wherein suspect sequences or customers are identified; these question marks represent uncertainty in the industry about who should be contacted if suspicious activity is identified.

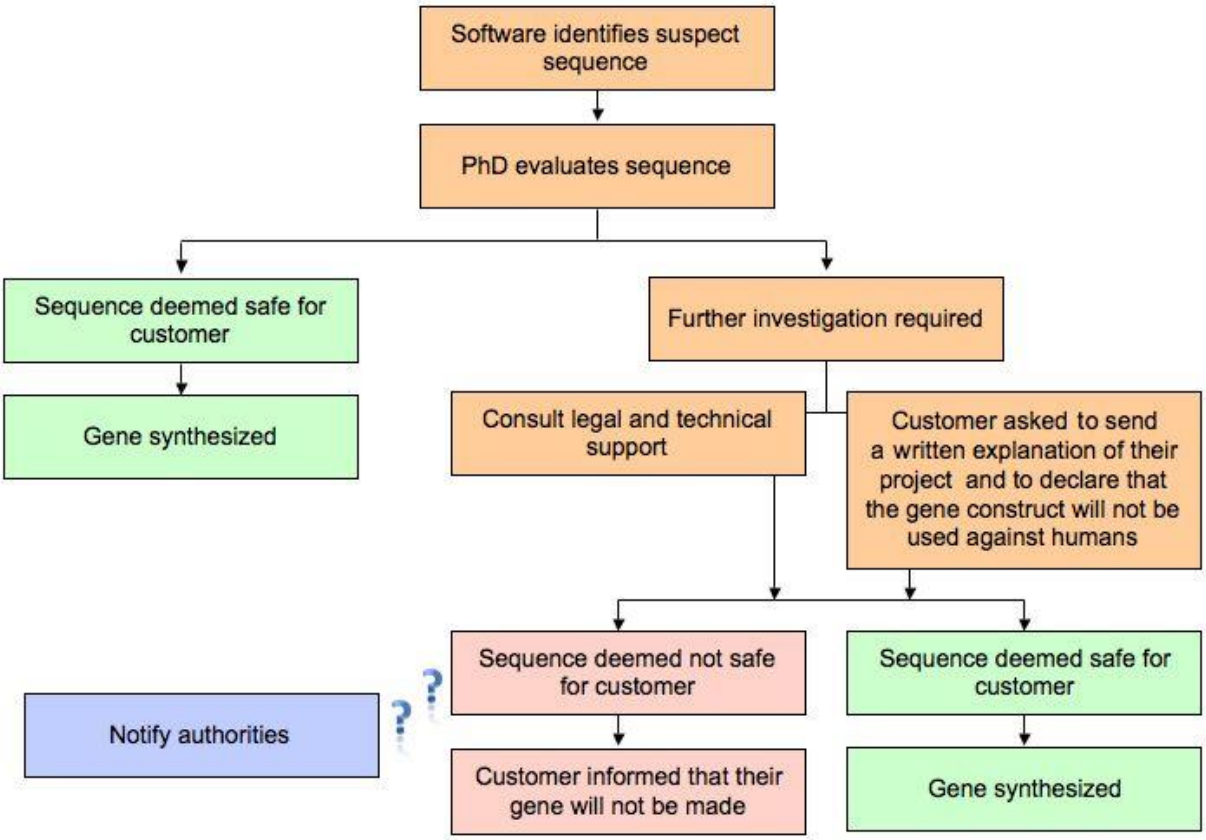


Figure 10. Flow chart of action taken by some companies when a sequence is identified as containing an element of concern. The question marks linking "Sequence deemed not safe" to "Notify authorities" represents the confusion in the industry about who to contact when a potential risk is identified.

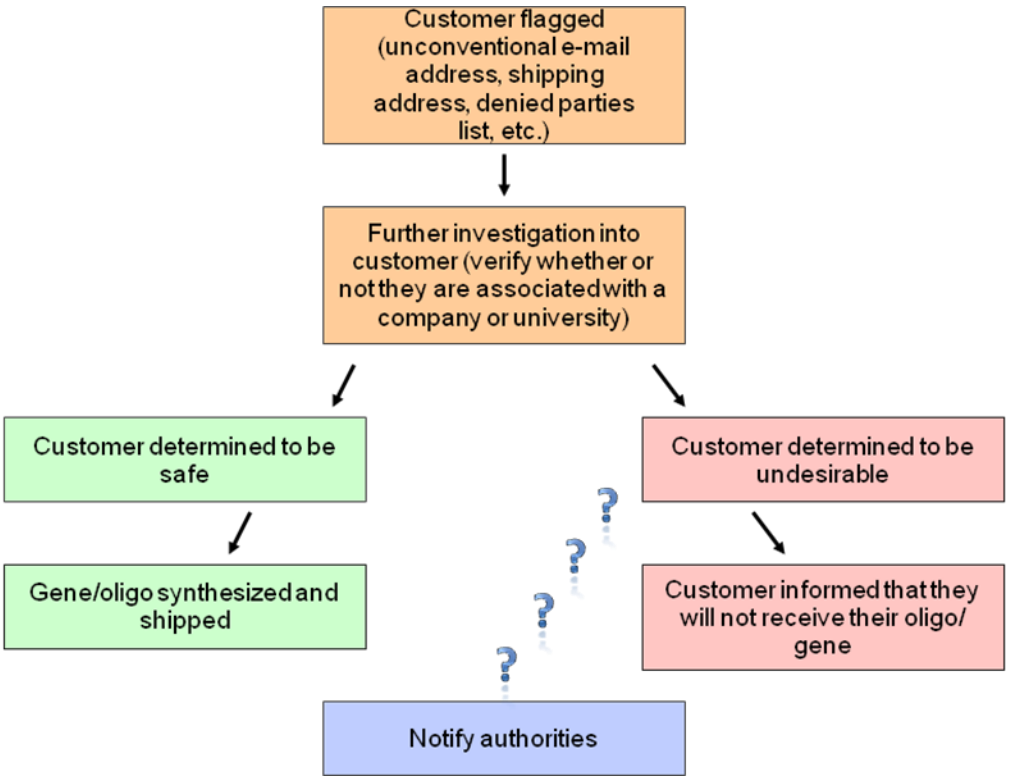


Figure 11. Flow chart depicting the steps taken by some companies when a customer is flagged due to an unconventional e-mail address, shipping address, or inclusion on the denied parties list. The question marks linking, "Customer determined to be undesirable" to "Notify authorities," represents the confusion in the industry about who to contact when a potential threat is identified.

Archiving Customer Information

As mentioned in Chapter 2, it would be advantageous to be able to track what sequences have been ordered and who placed the order in the event of misuse of CSNA technology. Every gene and oligo synthesis company we spoke with archives past orders. Many of the companies we spoke with indicated that order information (including sequence-specific data, customer names and addresses) are stored indefinitely; for the few companies that do not archive information indefinitely, information is stored for a minimum of two years. Although most companies feel releasing this information without a warrant would be in violation of their customers' confidentiality agreements, all agreed the information would be made available if needed for a forensic investigation. Furthermore, some suppliers mentioned that, although data is stored, it is not stored in a database that facilitates searching.

The Next 5 Years

The gene synthesis and oligo industries are undergoing a transformation. With adjustments in price, synthesis speed, fidelity, and demand for synthetic DNA products, it is important to understand how the industry will change in the coming years. Below we detail changes expected by synthetic gene and oligo providers themselves in the next five years. We found no useful data outside of our interviews that predicted likely changes in the industry beyond this year. Gene suppliers predict a different future for their industry than those in the oligo industry, although all agreed that both CSNA industries business will stay steady or grow over the next five years.

Synthetic Genes

Those in the gene industry anticipate an increase in demand for synthetic genes. Many in the synthetic gene industry pointed out that there is more to making a functional gene than simply synthesizing and assembling DNA, which is why some feel the industry may split into two components: those that synthesize and assemble genes, and those that specialize in gene design. Companies that specialize in gene design would move from synthesizing genes based on sequences submitted by the end-users to constructing genes with a slightly altered DNA sequence based on the protein the customer plans to express. Sequences would be optimized for codon usage and protein expression, while oligos used to produce such genes might be purchased from an outside provider.

Everyone we spoke with in the gene industry anticipated that in the next five years the prices for synthetic genes will decrease. As the technology becomes cheaper and additional uses for synthetic genes are discovered, all agree that the demand for genes will increase, although this increase was not quantified by any respondent. Likewise, there is a consensus among the industry that gene production speed will increase two- to three-fold over the next five years. Although there is not agreement about whether there will be an increased demand for larger sized genes than are currently being made, all do agree that if the demand increases the industry will be able to meet the customers' need.

Oligos

While gene companies anticipate a transformation in demand, price, and production speed, the oligo companies do not anticipate this same level of change. Some oligo industry representatives pointed out that in the last few years there has been some consolidation in the industry with more to be expected. As prices for oligos have decreased, many of the smaller oligo companies have been bought out by larger companies, or driven into specialized niches (such as modified oligos, RNA oligos, morpholinos or large scale synthesis). Our internet search for oligo providers identified several companies such as AnaGen Technologies, Bio Pioneer Inc., and Picoscript which have discontinued oligo manufacturing. Others such as e-oligos and DNA Technologies, which sell oligos, now outsource their synthesis. Even Qiagen, which according to the 2006 Bioinfomatics survey provided oligos to 13 percent of the survey respondents, no longer produces or sells custom oligos.[109]

Many of the oligo producers commented that the industry has already matured and production speed and fidelity have reached a plateau, therefore they do not anticipate changes in these areas in the coming years. Although a few in the industry think there will be a decrease in oligo price in the next five years, most do not think that oligo prices will drop given that oligo prices are already very low. Greater than 91 percent of the oligo companies we spoke with anticipate an increased demand for oligos in the coming years (the remaining feel the demand will stay about the same). While most feel the industry will see an increase in demand, there is no consensus about the types of products will be most needed. Some feel that increased demand will be for short oligos while others feel it is long oligos that will gain in popularity. Several companies mentioned that at least some of the increase in oligo need will stem for research currently being conducted for use of oligos as therapeutics; as this practice becomes more common the increased need for large quantities of (mg or kg quantities) may increase.

We found it interesting that interviewees from both the oligo and gene synthesis industries expect that demand with grow for their products. This sentiment indicates to us that either the industry is being overly sanguine or that any encroachment on the oligo market by the gene synthesis companies will be more than compensated by demand for oligos from those new to molecular biology.

[109] Oligos that used to be sold by Qiagen are now sold by Operon, which was the result of a management buyout of Qiagen's global oligo business in July 2004. Interestingly, Qiagen acquired Operon to capture its oligo business in 2000.

[Commercial proprietary information redacted.] Given that the predictions of industry jive with the limited trend data available we must assume at this point that the gene industry will *not* overtake the oligo industry in the next five years in production volume or income.

A Note About Amateur Biotechnologists

Another interesting factor, generally not addressed in the literature, is the potential rise of amateur biotechnologists not associated with easily identifiable institutions.[110] As the technology to create or imbue life with useful or amusing features spreads, becomes packaged into kits and becomes cheaper, amateur biotechnologists will emerge. This possibility invites the comparison between amateur biotechnologists and amateur computer programmers enabled by the advent of the personal computer. Amateur computer programmers were responsible for the invention of beneficial programs (such as operating systems, freeware and games) and also malicious programs (such as computer viruses). Similarly, amateur biotechnologists could produce great benefit and great harm. Any oversight system that prevents customers who are not affiliated with a research institution from receiving CSNA should be willing to bear the cost that a potentially very energizing segment of the population will be excluded from innovating using the tools of molecular biology. Also, a system that precludes these customers from gaining access to CSNA may financially hurt CSNA providers in the long term and potentially create a black market that could be exploited by the well-meaning and sinister alike. Conversely, a system that provides no oversight of amateur biotechnologists will be overlooking a potentially serious safety and security risk.

Conclusions

A review of the CSNA industry reveals that it is actually composed of two, easily separable, components: the gene synthesis companies and the oligo synthesis companies. The gene synthesis industry is relatively small, producing less than 0.2 percent of the volume of products delivered by the oligo industry each year. Many providers of synthetic genes are aware of the potential for product misuse and are therefore open to the idea of screening sequences for pathogenic elements; many are already doing so. Most in the oligo industry, however, do not feel that screening sequences is necessary or practical due to the small size of their product and the high-throughput nature of the industry. Despite the opposition of the industry, given the information provided by oligo manufactures, it appears there is a 15-60 minute window between the time orders are placed and synthesis begins where sequence screening could be introduced. Both synthetic gene and oligo manufacturers are amenable to the idea of screening customers to ensure products are only shipped to those with legitimate ties to the scientific community and many in both industries are already screening customers. Providers of both synthetic genes and oligos have expressed their desire for a designated contact within the government where suspicious customers or orders can be reported. With the CSNA industry already valued at over $1 billion and growing, representatives from the industry hope to prevent the misuse of technology without substantial costs to their business or customers.

The Customers

[Commercial proprietary information redacted.] The enormous customer community poses a daunting challenge to any oversight regime, especially when combined with the fact that the community is extremely active: according to our research about 30 million CSNA are ordered each year. Given the

[110] We recognize Drew Endy, Professor of Biology at MIT, for alerting us to this point.

quantity of stakeholders in an oversight regime that affects the CSNA industry, their opinions must be kept in mind when considering the options for such a regime.

It is interesting to consider how many of these scientists would order CSNA that may be associated with pathogenicity and, by extension, how many sequences of concern may be ordered each year. Although we could not determine how to arrive at a robust estimate for this quantity, we attempted to approach this number by drawing conclusions from the quantity of scientists working in various fields. Of the scientists listed as full members by the American Society of Microbiology (ASM), 19,000 are in divisions that suggest they work with pathogens and primarily work in the United States.[111] If the same proportion of microbiologists order CSNA as life scientists in general, then 11,500 microbiologists who work on pathogens will order CSNA.[112] Therefore, if we assume that these microbiologists order the same volume of CSNA as non-microbiologists, 5 percentof the total CSNA orders will be from microbiologists who work on pathogens. We could not determine what proportion of CSNA orders from these microbiologists would be based on sequences from pathogens or from elements that actually are associated with pathogenicity in the pathogens. Given that most researchers work with pathogens to study pathogenicity (those who study processes other than pathogenicity, such as microbial physiology, typically do so in non-pathogenic organisms, which are more convenient to work with simply because they are non-pathogenic), we can assume that most of their CSNA orders would be associated with pathogenicity. However, even if we assume that only 20 percent of CSNA orders from microbiologists who work on pathogens would match sequences from pathogens, still approximately 1,000 sequences would be ordered per day that would match sequences in pathogens (if we assume that European scientists follow the same trends as US scientists).

Furthermore, given that there are approximately 14,000 researchers registered under the Select Agent Rules, we can make some further estimates useful for our analysis. Assuming that 60 percent of these scientists order CSNA, we can estimate that roughly 8,400 scientists who work with CDC select agents are ordering CSNA.[113] Using the same logic as above, we can estimate that approximately 3 percent of orders will be from scientists working with select agents. Even if only 20 percent of these orders will be based on sequences from the Select Agents themselves, several hundred CSNA sequences per day that match Select Agents will be ordered.

Needs and Constraints of Customers

Although we have data from [Commercial proprietary information redacted.], we had no data on the impact of potential changes within the industry as it would affect the buying patterns of its customers.

To gather these data, we conducted interviews with researchers in industry and academe. The short time frame of this project limited the quantity of interviews we could reasonably accomplish, so we strove to capture a wide variety of opinion with as few interviews as possible. The academics we interviewed were from all career stages: graduate student, post-doctoral fellow, research fellow and principal investigator; from large, prestigious research institutes and small colleges. The industrial customers covered biotechnology companies, pharmaceutical companies and diagnostics companies (and one not-for-profit

[111] We obtained data on the full members of ASM by division and region via a request to the ASM, July 23, 2007. We note that although their acronym would suggest that ASM members are in the US, about a third of their members are, in fact, foreign. We considered only the US members for this analysis. We conceded that many microbiologists are not full members of the ASM, so this may be an underestimate.

[112] We assume that this number is an underestimate of the true proportion of microbiologists who order CSNA because microbes are less amenable to non-molecular assays than macroscopic organisms.

[113] It is important to note that not all researchers registered under the Select Agent Rules actually work with the pathogens. However, this estimate counterbalances the fact that there are many researchers who do not work with the organisms themselves and yet may work with some of their molecular components.

biomedical research institute). The companies interviewed ranged from large, established companies to small firms still funded by venture capitalists.

The interviewees in academe and industry reported that they used oligos for a variety of reasons, including direct cloning using genes, sequencing, mutagenesis, capture and signaling reagents for diagnostics, RNA interference and generic PCR. Even though we executed slightly less than 20 interviews, some trends in the data regarding requirements for their oligo suppliers were clearly apparent.

Academe

Proprietary nature of oligos. Only one interviewee thought that the sequences he ordered were proprietary. This researcher works in a competitive field and he felt that a competitor could determine which genes he is working on by scrutinizing his orders. Other researchers considered that their oligos would not provide useful information to a competitor. All trusted the government to keep their data in confidence.

Choice of producer. Only one researcher interviewed uses a core oligo synthesis facility, and this researcher cited convenience as the main driver for using this facility even though he recognized that the core facility was not cost-competitive. Some interviewees had access to a core oligo synthesis facility but rarely used it because these oligos were more costly than those made by private companies and the interviewees considered the core facility oligos to be generally of poor quality.

All researchers interviewed use only one provider of oligos, except when modified oligos were needed, in which case a specialty producer was used.

Cost was cited most often (by 70 percent of respondents) as a factor important for choosing a particular supplier. Two researchers remarked that a 25 percent increase in cost would cause them to search for another supplier. Timeliness was a concern; half of the interviewees cited it as an important factor. Some mentioned that an extra day delay would be acceptable if the supplier was still the cheapest and most reliable supplier. Others mentioned that a delay of an extra day would be a "deal breaker". Time seems to be less important if a customer is ordering many oligos at once (because they may be gearing up for a large project), but critically important if only a few oligos are ordered at a time (because these oligos are likely to be the critical reagents in the next step of an ongoing project—like sequencing primers). Also, time is less critical when ordering long oligos or genes as these orders are processed over the course of weeks.

Three interviewees mentioned quality (accuracy and purity) as an important factor. Two interviewees each mentioned customer support and history with a supplier as critical factors in choosing a supplier. The only user of a core facility mentioned convenience as his most important factor (even though he acknowledges that the core facility is very expensive).

Industry

Proprietary nature of oligos. Surprisingly, about half of the industrial customers interviewed suggested that they considered no information regarding their oligo orders proprietary. For those that considered the orders proprietary, all considered the specific oligo sequences proprietary, not the volume or frequency of orders. Furthermore, there is a great degree of trust that the oligo industry will protect the proprietary information of their clients because the industry is perceived to have the appropriate procedural and IT systems in place to protect the information. Also, some interviewees commented that the sheer volume of orders (some of these customers order more than 500 different oligos at a time) that oligo companies receive from large biotech and pharmaceutical companies would complicate the identification of the

sequences that would be truly interesting to a competitor. For large companies, oligo orders that are considered extremely competition-sensitive are made in-house to avoid any chance that information would be leaked by a second party.

Choice of producer. Of the industrial customers interviewed, only two still perform any synthesis in-house. These companies make oligos in-house do so because outside companies have not made specialized products (branched DNA, specifically) with the desired quality or because they did not wish to chance the leak of proprietary orders. Interestingly, more than half of those interviewed produced oligos in-house in the past, but chose to exclusively outsource the production of oligos. Primarily, the reason for outsourcing was efficiency (of personnel and capital equipment) and quality control. Even companies that use oligos as part of their core products (diagnostic companies) outsource the synthesis of the oligos, though they order the oligos in a larger quantity than is typical. For the groups interviewed, this change to outsourcing oligo synthesis occurred between one and two years ago.

Oligo producers are typically identified by industrial customers via word of mouth. For the larger companies, final choice of supplier is preceded by face-to-face meetings and trial orders that are closely inspected for accuracy and purity. The choice of outside producer was primarily influenced by the perceived "quality" of the product, even for small companies. Timely delivery of the oligos was also considered important; industrial customers stated that two day delivery of short oligo orders and delivery on the promised date for longer oligos is expected. Furthermore, good quality control, availability of technical staff for troubleshooting, flexibility and ability to handle "unusual" orders were considered important. All of these considerations were ranked as more important than price.

Generally, industrial customers work with one oligo supplier for most of their oligo needs. Customers will order from other oligo companies if unusual formats, modifications, or DNA analogs are needed. Additionally there seems to be substantial customer loyalty amongst industry customers.

Attitudes of Customers towards Oversight of the Industry

In this section, we provide some details on the opinions of the customers of the CSNA industry on the potential oversight of the industry. [Commercial proprietary information redacted along with Figures 12-18.]

Conclusions

Even though academic and industrial customers tend to use only one supplier of oligos, a regulatory regime would hurt the oligo synthesis industry if it imposed a delay of as little of a day on the delivery of oligos (not genes). From the standpoint of academic customers an oversight regime that increased the cost of oligos by as little as 25 percent may also cause these customers to look for cheaper suppliers. The majority of academic customers would not mind if data on their oligo orders left the initial supplier for scrutiny elsewhere, provided that this transfer did not result in a delivery delay for their oligos, because they did not feel that their oligo orders contained proprietary information. Some industrial clients, however, thought that their data was proprietary and wanted all possible controls in place to reduce the chance that their data would be leaked.

Furthermore, given that academic and industrial customers tended to use a single supplier, changing to new suppliers happens rarely. Therefore, a synthesis company will rarely encounter an order from a new customer, facilitating the screening of only new customers.

If a screening system was to address the common reasons behind resistance to oversight mentioned by these interviewees, it must effectively prevent easy circumvention of its efforts yet be flexible or efficient enough to not impose a burden on suppliers and researchers. Addressing the primary concern that terrorists may be able to circumvent any screening system will ultimately require oversight of other activities, such as the purchase of synthesizers and synthesis reagents and registration of existing synthesizers, and require the system to encompass CSNA providers globally. The secondary concern of implementing a screening system is that it would burden the industry and researchers with increased costs and turnaround-time. Requiring customers to submit their credentials and a description explaining why they need a suspicious oligo with their order may help reduce the amount of additional turnaround-time by cutting out the need for a third party to uncover this information. Furthermore, if companies were able to store customer credentials in a database, it would eliminate the need to have customers resubmit their credentials with each subsequent order. It will certainly take time, effort, and the collaboration of researchers, industry, and agencies in order to develop an effective screening system palatable to CSNA customers.

Review of CSNA Screening Tools

Background

Similar regions of DNA, RNA or protein are identified through a process called sequence alignment. A very simple alignment could be done by hand. Two sequences would be compared side by side with some simple metric applied to nucleotides that are identical, those that have conservative differences and those that are very different. To compare a sequence to many other sequences, as in a database, an algorithm is needed. These generally work in two ways. Local alignments identify regions that are similar within longer sequences that are not alike. Global alignment considers all nucleotide or protein similarity over the entire length of the sequence.

The Needleman-Wunsch[114] algorithm is a general global alignment technique. This is an iterative algorithm that produces a maximum match (e.g. the largest number of matching nucleotides). It assigns similarity scores in a 2-way array and computes the score of all pathways back to the beginning of the

[114]Neeleman, S and Wunsch, C. "A general method applicable to the search for similarities in the amino acid sequence of two proteins", *Journal of Molecular Biology*. 48(3):443-53. 1970.

sequence. Then it gives a score to the most likely alignment. The Smith-Waterman[115] algorithm is a general local alignment method that has been further developed by Karlin and Altschul.[116] Unlike Needleman-Wunsch, it examines only local areas, not the sequence in its entirety. This method may not be suitable for identifying exact matches but is important for tracking functionally important local sequences. When sequences are mostly similar, global and local algorithms will produce similar results.

When comparing multiple sequences and/or large databases, sequence alignment methods are time and memory intensive. Faster computers and better programming have allowed for the development of algorithms that make the process straightforward for the user. The two most common open access platforms for similarity searching are FastA (Fast Alignment) and BLAST (Basic Local Alignment Search Tool).[117,118] Each uses approximations of the Smith-Waterman and Needleman-Wunsch algorithms to search databases of known sequences. The database and algorithms used in the comparison can be selected by the user to tailor each search.

FastA (Fast Alignment) was the first widely used algorithm for database similarity searching.[119] The program looks for local alignments by scanning sequences for small matching segments. BLAST (Basic Local Alignment Search Tool) is a set of algorithms that finds local alignments as well as global similarity between two sequences of nucleotides. BLAST finds alignments by identifying pairs of words (nucleotides chains) that match. It then extends those words in both directions to give a maximal sequence pair (MST). BLAST determines statistical significance by comparing the similarity found to that which would be expected by chance (as determined by simulation).

The BLAST[120] algorithm is a heuristic search method that seeks words of length W (default = 3 in standard BLAST) that score at least T[121] when aligned with the query and scored with a substitution matrix. Words in the database that score T or greater are extended in both directions in an attempt to find a locally optimal ungapped[122] alignment or HSP (high scoring pair) with a score of at least S or an E value[123] lower than the specified threshold. HSPs that meet these criteria will be reported by BLAST,

---

[115]Smith TF and Waterman MS. "Identification of Common Molecular Subsequences." *Journal of Molecular Biology* 147: 195-197. 1981.

[116]Karlin, S and Altschul, SF. "Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes" *PNAS* **1190** (87); 2264-2268. 1990.

[117]Altschul, SF, W Gish, W Miller, EW Myers, and DJ Lipman. Basic local alignment search tool. J Mol Biol 215(3):403-10, 1990.

[118]Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z.Zhang, W. Miller, and D. Lipman. Gapped and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25(17): 3389-402. 1997.

[119]Lipman, D.J. and W.R. Pearson. Rapid and sensitive protein similarity searches. *Science*; 227(4693) 1435-1441.

[120] BLAST is available online through the National Center for Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/BLAST_algorithm.html

[121]Optimal score averaged over a set of random sequences.

[122]Gaps are missing nucleotides in a sequence

[123]E (Expect) value. The number of different alignments with scores equivalent to or better than S that are expected to occur in a database search by chance. The significance of a match is reported as an e-value or p-value and empirically derived by comparisons of random sequences of the same length and composition as the sequence to sequences in the database. The lower the E value the more significant the score. It is a reflection of both the size of database and the scoring system in use.

provided they do not exceed the cutoff value specified for number of descriptions and/or alignments to report.

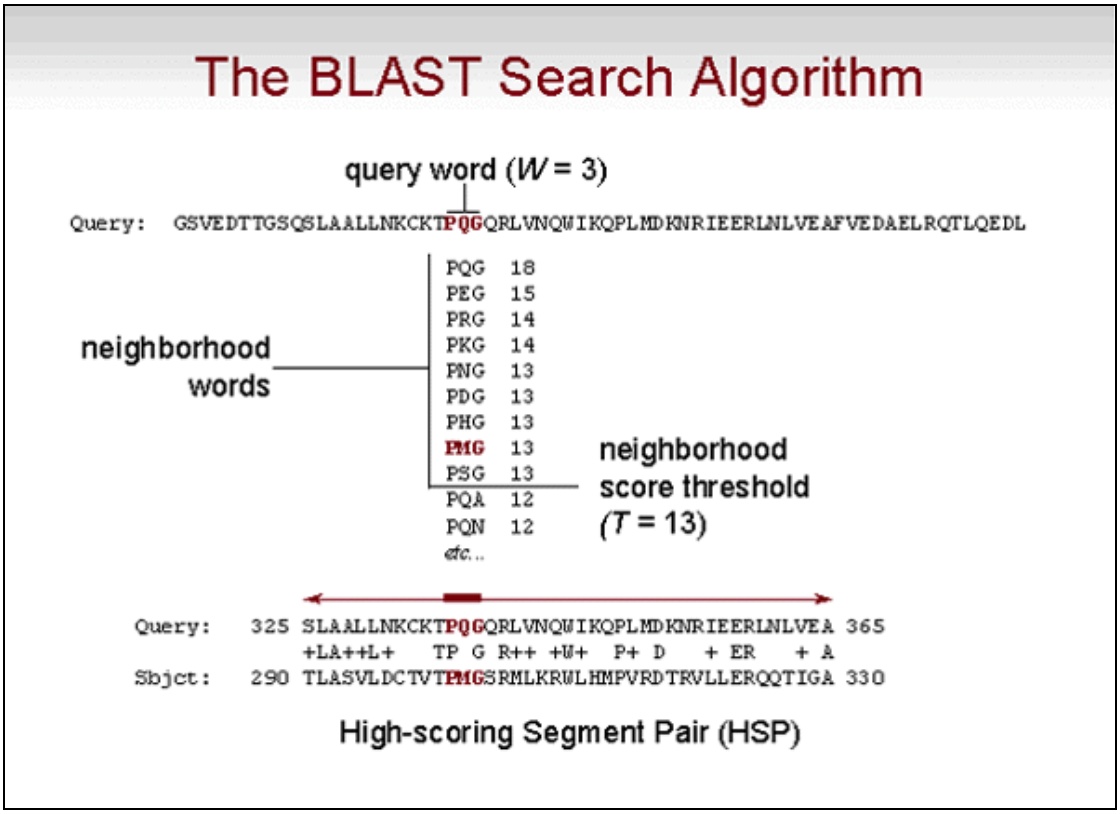


Figure 19. The BLAST search algorithm, described more fully in the text above.[124]

Within both FastA and BLAST, the similarity of nucleotide bases in a DNA sequence is measured using a substitution matrix. Substitution matrices give a score for each match or exchange that occurs for each nucleotide. The matrix chosen for any analysis will determine the conclusion a comparison reaches. These models were developed to determine homology and find similarities in proteins. Each model assumes a certain representation of evolution. The two most commonly used substitution matrices are BLOSUM and PAM. They work on opposing scales with higher levels of PAM being more dissimilar and higher levels of BLOSUM being more similar.

Additionally, BLOSUM (Blocks Substitution Matrix) a set of matrices that analyze protein similarity in larger blocks,[125] uses conserved blocks of sequences that are assumed to be of functional importance. While the PAM matrices are based on global alignments, BLOSUM uses only those areas of a sequence that are highly conserved. This method clusters similar sequences so that each counts only fractionally to the total tally. These matrices can be set to return results that meet a probability threshold. In general, PAM is best suited for sequences that are similar while BLOSUM is best suited for sequences that are more distantly related.[126]

---

[124]Figure Taken from NCBI: BLAST tutorial. http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/BLAST_algorithm.html
[125]Henikoff, S and Henikoff, JG. "Amino Acid substitution matrices from protein blocks" *PNAS* 89: 10915-10919.1992
[126]National Center for Biotechnology Information, National Library of Medicine, "Scoring Systems." Available at: www.ncbi.nlm.nih.gov/education/blastinfo/scoring2.html.

BLAST has a specific algorithm for matching sequences of nucleotides, BLASTn. BLASTn compares a sequence against a nucleotide sequence database indexed into possible words and is intended to find short, nearly exact matches (identity over homology). This algorithm uses a simple similarity matrix that uses a score of 5 for matches and -4 for non-matches[127]. The default word size is 11 and can be increased to increase speed and limit hits or reduce to increase sensitivity.

Example: BLASTn matrix (similarity)

	A	T	C	G
A	5	-4	-4	-4
T	-4	5	-4	-4
C	-4	-4	5	-4
G	-4	-4	-4	5

Within this framework, there are anomalies within sequences that can significantly change the matching score. Gaps are introduced to compensate for insertions or deletions in a sequence. Gap penalties[128] are given for the occurrence of each gap with a small additional penalty for the size of the gap. Naturally, nucleotide base changes, or mutations, occur within genes over time, which can be divided into transitions and transversions. Transitions conserve the number of rings in nucleotide base, whereas transversions do not and a similarity matrix can weight transitions and transversions differently. This weighting can improve recognition of some distantly related, or somewhat differing, sequences. The most important nucleotide change in terms or screening occurs in the wobble position. The wobble position refers to the third position in the three nucleotide code (called a codon) for an amino acid. Many amino acids can be encoded by multiple codons that differ in the nucleotide in the third position. For instance, the sequences CUU, CUC, CUA, and CUG all code for the amino acid leucine. Therefore, a sequence need not match exactly to code for the same amino acids and ultimately have the same functions, so these changes occur frequently even in conserved genes.

In this section, we summarize the information gathered in interviews with CSNA customers and suppliers that speaks to their requirements of an ideal sequence screening tool. These requirements are underpinned by the normal business practices of the industry and the demands of the consumers, and, importantly, not driven by security needs. Security needs for a screening tool can only be identified through a thorough process risk assessment. It is possible that security needs may conflict with the requirements derived from the needs of the marketplace.

Requirements of Customers and Suppliers of Sequence Screening Tools

If a screening tool is to be used to screen oligos, it should give an actionable answer in a few minutes or less. This time frame is generally less than the time that an order waits after it is received but before processing occurs. If screening can be done in this time frame, a company will not have to produce orders only to throw them away. In contrast, the process of designing a gene takes longer than even inefficient sequence screening processes so the time for screening is not a critical parameter when screening genes. (However, more rapid screening is desirable even for gene orders as the determination that an order cannot be processed late in the design stage wastes the time of researchers at the gene suppliers.)

Cleary, the false hit rate should be as low as possible; a parameter that will be driven by tool performance and the desired false miss rate, which must be determined by a thorough risk assessment. However, we can provide some cost estimates of various false hit rates. For the gene industry, a primary screening tool with a false hit rate of one in a thousand would result in roughly 50 false hits a year. Even if we assume

[127] States, DJ et al. "Improved sensitivity of nucleic acid database searches using application-specific scoring matrices" 1991 Methods: A Companion to Methods in Enzymology.Vol. 3 pp. 66-70. 1991. Available at: http://blast.wustl.edu/doc/ntmats.pdf,

[128] Gaps receive a negative score in the scoring matrix giving a lower total score in the sequence alignment. Gaps indicate that the sequences are less well matched than the same sequence would be without gaps.

that one hour of follow up is required to resolve false hits, this rate would require a full-time person just more than a week to resolve all false hits in a year system-wide. This same false hit rate in the oligo industry would result in roughly 25,000 false hits a year. Once again if we assume one hour to resolve false hits, slightly more than 10 FTEs must work for a year to resolve false hits. In contrast, if we use the statistical false hit rate of BlackWatch, a screening tool currently in use (one in 10,000), only 2,500 false hits would result, requiring about one FTE full time to resolve false hits system-wide.

Ideally, the screening tool should be able to automatically take sequence data from the processing machinery so that no additional work is required to enter the sequences. If this is not possible, operators should be able to enter sequence data by "cutting and pasting". A system that requires the operator to enter sequences by hand will take too much time to be usable for the millions of orders processed by the oligo industry and be impossible to use in the gene industry because the products are too long to reliably be entered by hand.

Clearly, the sequence screening tool should give unambiguous results. An output such as "no match detected" may be desirable in the case of misses. In the case that a query sequence matches a sequence of concern, the system should let the user know which pathogen it matched. This additional information will greatly help the user in confirmatory screening.

CSNA suppliers were very steadfast in their requirement that data on orders should not leave the company. In order to comply with this requirement, the sequence screening tool (which includes the interface, algorithms and sequence database) should be downloadable onto computers at the screener's facility. In this way, query sequences can be entered into the tool without any information being sent from the company.

As you will see, one tool, called BlackWatch, currently in use by some gene suppliers, satisfies many of these requirements. It is downloadable onto the computers of screeners obviating the need to ship potentially proprietary data outside a firewall, it has a statistical false hit rate of one in 10,000 and gives unambiguous results for a hit or a miss. Furthermore, the user is given information on which sequence a query matches, facilitating confirmatory screening. Additionally, the tool works quickly, giving an answer in typically less than a minute.

Existing Sequence Screening Tools

The Smith-Waterman algorithm is available as part of the freely available search program, FastA[129], as well as the commercial packages Cray XD1 and GenCore6 by Bioacceleration and DeCypher and CodeQuest by TimeLogic. The set of BLAST[130] algorithms is available through the National Center for Biotechnology Information (NCBI). They are linked with the full set of all the genes, nucleotides and DNA entered into the databases in GeneBank. In addition, there are many free and proprietary search systems that are based on BLAST search algorithms. For example, GenomeQuest Inc. offers an eponymous commercial program that incorporates all new patented sequences into BLAST searched databases.[131]

---

[129] FastA is available from multiple open access websites including: FastA Sequence Comparison at the University of Virginia (http://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml), and European Molecular Biology Laboratory-European Bioniformatics Institute (EMBL-EBI): FastA Protein Similarity Search (http://www.ebi.ac.uk/fasta33/)

[130] NCBI-BLAST homepage (http://www.incogen.com/public_documents/vibe/details/NcbiBlastn.html). BLAST is also available through the EMBL-EBI (http://www.ebi.ac.uk/blast2/) and Kyoto Bioinformatics Institute (http://blast.genome.jp/).

[131] GenomeQuest Inc. Homepage. http://www.genomequest.com/index.html

A commonly used program for sequence screening, Craic Computing's BlackWatch,[132] automatically uses the appropriate BLAST algorithm depending upon the sequence length. BlackWatch searches against nucleotide and protein databases compiled from the CDC, USDA, and Department of Commerce regulations on hazardous biological agents.[133,134] As new sequences are added to GenBank[135] they are incorporated into BlackWatch updates. Entelechon has recently developed pathoGENEDetective, which also uses BLAST, to allow online screening of sequences with a customizable set of pathogen sequences and common similarity detection systems.[136] While BlackWatch, minimizes false positives with stringent matching criteria, pathoGENEDetective focuses on database management. That is, it only screens against genes that are deemed truly virulent, eliminating positive hits from organisms that are similar to pathogens in many non-virulence genes.

Before continuing our description, we must take a step back. In any discussion of CSNA sequence screening, it is important to note that the currently available sequence alignment algorithms were not designed for the identification of any DNA sequence. They were designed to investigate homology[137] and find similar functional groups in disparate organisms. These algorithms are not straightforward and were designed as tools for researchers that are at least somewhat acquainted with the sequences they are querying. One of the most common methods of screening, BLAST, is a heuristic algorithm (that is, it is designed to make tradeoffs in completeness for speed and seeks the easiest path to finding a match). In its default settings, it emphasizes speed over sensitivity.

The number of false-hits and false-misses in a search will depend on the settings of each individual search. Furthermore, any simulation used to test the rate of false results will depend on the parameters used in the simulation.[138] BLAST uses randomly generated sequences of the same length and composition as the query to compute a probability distribution and thus give a significance value[139] to each sequence matching score. While this value is analogous to statistical significance, it does not have the same associated parameters and confidence intervals. Thus the cutoff values (for instance, the maximum accepted e-value) to give a positive hit will be key in determining the accuracy of any BLAST-based screening process.

$$E = Kmne^{-\lambda S}$$

Where E is the expected number of sequences to have a score (S), n is the length of the query sequence and m is the length of the database sequence, and K and λ are the parameters of the search scale and scoring system.

The default threshold e-value for a standard BLAST search is 10. This means it returns only those matches that have an e-value less than 10, which indicates that 10 matches with that score are expected to be found by chance. This threshold is intended to be reasonable when searching for homology, but it is

---

[132] Cray XD1 Smith-Waterman brochure: http://www.cray.com/downloads/SmithWaterman.pdf

[133] Craic BlackWatch®, Hazardous Biological Agent Sequence Detection. https://biotech.craic.com/blackwatch/regulations/regulations.html

[134] Referred to in this section as controlled pathogens.

[135] GenBank® is the National Institute of Health's database of all publicly available DNA sequence. NCBI GenBank overview. http://www.ncbi.nlm.nih.gov/Genbank/index.html

[136] Fischer, M T. Blank, P Eser et al., Towards an automated screening of biorisk-associated DNA and protein sequences.2007 Synthetic Biology3.0 poster. www.p-detective.org

[137] Similarity attributed to descent from a common ancestor (NCBI BLAST terms: http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/glossary2.html)

[138] Simulations could use randomly generated sequences or they could use some experimental model that incorporates the typical frequency and or order of nucleotides.

[139] E-value (expect value) – the number of random sequences expected to have score higher than the current search.

almost certainly too high for screening purposes. If a company screens gene sequences by manually entering sequences into the NCBI online BLAST algorithm,[140] that company would have to determine what threshold to set as a positive hit as well as the other parameters of the search.[141]

BlackWatch uses the e-value, total match score and span to determine a match. Span refers to the portion of the query sequence that matches in the database. A query sequence with only a partial match against the database would have a lower span than a sequence that matched the database entirely. BlackWatch sets the e-value threshold at 1×10^{-4} meaning that only 0.0001 matches with that score are expected to be found by chance. This is considered to be a conservative threshold that will signal a hit only when the sequence is a close match. The developers of this, or any other screening software, have the ultimate authority on what sequences will or will not give a positive hit when using their tool.

Ultimately, any sequence screening would be placed within an overall screening protocol. Kahn[142] proposed incorporating sequence screening against a pathogen database with a virulence[143] weighting procedure and some sort of threat determination. In this system, virulence is determined by the degree of pathogenicity of the organism as well as the data known about the specific gene that a query sequence matches. The threat refers to the likelihood of an oligo being used nefariously, including the proportion of the organism covered, the number or hits on this organism over time, the country of the order and the number or users[144]ordering an oligo from one organism. To fully assess cross-order correlation, all screened orders would have to go through some sort of clearinghouse where all oligos are screened. Finally, this system gives a risk score as a probability range (threat likelihood x virulence = risk). It would be up to the supplier to determine if each individual order was for a legitimate purpose.

Testing Screening Systems

We explored the practicalities of screening and false hits and misses by running multiple sequences through both the online NCBI BLAST algorithm and BlackWatch software. We selected 13 pathogens from the controlled agents list[145] and then selected three gene sequences from each. Beginning with 20 nucleotides from each gene, each sequence was put into BLASTn and BlackWatch and the outputs recorded as True or False and Positive/Negative.[146] If the screen did not identify the sequence, we added nucleotides from the gene sequence until the correct microbe was identified by BLAST and BlackWatch. We then modified each sequence by first changing nucleotides in the sequence itself and then by adding six incorrect random nucleotides to the end of each sequence (which simulates the addition of a restriction endonuclease site for cloning. All sequences screened were recognized as true hits in BLASTn at 35 nucleotides or less. The average length for recognition as a true positive was 20.7nt. BlackWatch, however, did not recognize short sequences as well. Only about 87 percent (27 out of 31) of the sequences screened were recognized as true positives at 35 nucleotides or less and 13 percent (4 out of 31) of the sequences were recognized at more than 35 nucleotides in BlackWatch. Interestingly, 9 sequences were not recognized by BlackWatch, presumably because they were not included in the database (Ebola, *C.*

[140]This approach is likely only feasible for companies that fabricate longer gene sequences. It would be overwhelmingly difficult to screen all small oligo sequences by entering them into a BLAST search manually.

[141]Such as algorithm (BLAST, BLASTn etc.), substitution matrix, and database.

[142]Kahn, A. "OligoSleuth: A Biological Database 'Sniffer' for Detecting Clandestine Bio-warfare Operations and Determining Possible Biological Defense Drug Targets" Conference poster presented at: Biodefense: research, technology and applications 2002.

[143]Virulence is defined here as: the degree or capability of causing disease by breaking down protective mechanisms in a host, or the potential harm that could be caused.

[144] It is assumed that someone constructing an organism for a biological attack might try to use several aliases and/or oligo suppliers.

[145]Compiled from the CDC, USDA, and Department of Commerce regulations on hazardous biological agents.

[146] In this exercise, we used only known pathogens, so there were no True Negatives.

Table 1. Sequence Screening Exercise: BLAST and BlackWatch									
Pathogen	**Sequence**	**BLAST**				**BlackWatch**			
		20nt	**>20nt**	**6nt added***	**1nt sub.****	**20nt**	**>20nt**	**6nt added***	**1nt sub.****
B. anthracis	agcgtgctggaaagaccaccttgtaattc		X	X	X		X		
	Ccatttgcacatattcgccc	X		X			X		
	Aattgctccaataactcatt	X		X			X		
C. botulinum	Cctattattttattcacttc	X		X	X				
	ttacggctgaatcattatgttttttag		X	X					
	Ctgtggagaagatcctgttc		X	X					
	Attgtttatacatttttttcttcaaaatatataaa agcaattaataaagctgtaga	X					X		
Ebola virus	Tattgttaaaggacagcatt	X		X	X				
	Ctaacaagatgacaactaga	X		X	X				
	tctccctgcgtgataatcaa	X		X	X				
F. tularensis	cctgagaaaatgctagaaaa	X		X	X				
	tgaaagcctcgattactctt	X		X	X		X		
	agtgtttattaccatagttt	X		X			X		
Guanarito virus	cttgggtggacaattaaggg	X		X		X			
	attctggactgttccccagt	X		X		X			
	gagattgagtccagcggaag	X		X		X			
Junin virus	tgtaactgttttctgtttgg	X		X		X			
	ttgtgatctagaaccaatat	X		X		X			
	aaactgtgtttcgttgaaca	X		X		X			
Lassa virus	tgggttcatgtgtcctactg	X		X	X	X			
	atttttttcaacagtctcctt	X		X		X			
	gtcgacactcttcaggtctt	X		X		X			

X – indicates a correct organism was identified in the first position of a BLAST search or a Positive Hit in BlackWatch
**6 random (incorrect) nucleotides added to the end of sequence*
***1 random (incorrect) nucleotide substituted within the sequence*

Table 1. Sequence Screening Exercise: BLAST and BlackWatch									
Pathogen	**Sequence**	**BLAST**				**BlackWatch**			
		20nt	**>20nt**	**6nt added***	**1nt sub.****	**20nt**	**>20nt**	**6nt added***	**1nt sub.****
Machupo virus	tcatgaaggagtatgatgta	X		X	X	X			
	gcctgtcctttgcctcataa	X		X		X			
	ttttcatagacatgagccta	X		X		X			
Marburg virus	caatgttggcgaacaatatc	X		X	X				
	ctccagaagacagaagacgt[147]	X		X					
	aggagttgcagttttagata	X		X	X				
Rabies virus	gatggaaaatcgcccaaccc	X		X	X	X			
	tcaatgatttctctctggtt	X		X		X			
	caagagtgagatgcagagag	X		X		X			
Rift Valley fever virus	tgatttgcagagtggtcgtc	X		X	X	X			
	gtgatgatgatggatttgtt	X		X		X			
	tagtgtttgtatctctaggg	X		X		X			
Variola virus	ctattaacttattagcgttg	X		X		X			
	ttatctgatgaaagataaactag		X	X	X		X		
	catagaaaatgattcacaatttat		X	X	X		X		
Y. pestis	tcacgataatcccctaatgc	X		X	X	X			
	gctattactttcgacccgct	X		X	X	X			
	acccattctcccacattgga	X		X		X			

X – indicates a correct organism was identified in the first position of a BLAST search or a Positive Hit in BlackWatch
**6 random (incorrect) nucleotides added to the end of sequence*
***1 random (incorrect) nucleotide substituted within the sequence*

[147]When screened through BLASTn, the second sequence of Marburg virus was identified in the third position after synthetic construct glycoprotein fusion protein pRAd and synthetic construct glycoprotein fusion protein pLAd. However, the E-values for all three positions were identical. It is likely that the two synthetic construct glycoprotein fusion proteins came from Marburg virus.

botulinum, and Marburg). While the results are neither exhaustive nor statistically conclusive, they do illustrate the issues that may arise as sequences are screened.

Random sequences from pathogens generally return the correct[148] microbe as a first match with 20nt in NCBI's BLASTn. The correct microbe is identified by BlackWatch when the query sequences are at least 20-25 nucleotides long. There are instances where BLAST e-values (the default sorting mechanism) are identical for many microbes and then they are listed alphabetically. If the correct microbe was in this list, we considered this a true positive.[149] As we were not performing enough tests to assess a false hit rate, we did not track instances where microbes other than the pathogen itself (or a very close relative) were in the first three records. Changing one nucleotide did not change the results of BLAST, but few sequences were still positive hits in BlackWatch for the shortest queries. Two nucleotide changes resulted in false negatives always in BlackWatch and roughly half the time in BLAST. Due to the nature of false negatives, without a significant automated testing scheme, it is not feasible to test enough sequences in order to fully evaluate either system to determine the likely rate of false negatives.

NCBI's BLASTn and BlackWatch differed most noticeably when six random nucleotides were added to the end of each sequence. This test simulates an oligo with a restriction site added to the end (for the purpose of cloning, for example), a common change from a native sequence. A BLAST search with query sequences with six additional nucleotides still identified the correct microbe as often as a search with the original sequence (albeit with lower scores and higher e-values). BlackWatch did not return a hit on any sequence with six added nucleotides. This is most likely due to the span threshold not being met in BlackWatch's settings.

The similar performance of BlackWatch and NCBI's BLASTn is to be expected. They are, in their mechanisms, identical since BlackWatch uses BLAST[150] as its searching algorithm. The key differences are the databases being searched[151] and the threshold settings. This brings us again to the conclusion that the setting of those thresholds will be paramount in determining the rate of false hits and false misses that result from any screening program. Because it is designed to be as user friendly as possible, BlackWatch has been set so that it's strengths are in identifying longer gene length matches, and to minimize false hits with short sequence queries.

In order to estimate possible false hit rates, we can make some assumptions about how a screening system might work. In the BlackWatch system, the e-value cut off (0.0001) represents a statistical estimate that there is a one in 10,000 chance that the match observed is serendipitous (i.e that the two sequences are similar but not related evolutionarily). If we ran 100,000 sequences a day for 300 days a year we might expect 3000 random sequences to match the database and give us a false hit. This represents a statistical upper limit as some positive hits for random matches would be discarded due to the score and span restrictions present within the software. However, as described in the section describing the customers of the CSNA industry, above, we can assume that about a thousand CSNA orders a day will be based on sequences from pathogens and several hundred CSNA orders a day will be based on sequences from select agents. Therefore, although only a few hundred hits could be expected by chance, the hits generated by sequences ordered by legitimate researchers would predominate. Additionally, as the sequence queries grow in length, they are expected to have fewer false positives for any given database.

---

[148]That is, the first organism listed in the BLAST results matches the known origin of the sequence.

[149]This may be unrealistic in a real-world screening scenario. If only the closest match is recorded, it may not recognize a pathogen that would result in follow-up.

[150]It chooses the proper BLAST search for the sequence size, in this exercise, BLASTn.

[151]NCBI's BLAST searches the entire GenBank nucleotide database. BlackWatch searches only nucleotides from controlled pathogens.

We assume that if both sequences and customers are screened, there will be some mechanism to determine which customers should legitimately get the CSNA orders that are associated with pathogenicity. Of greater concern is the possibility that a sequence accidentally matches one of interest. Using simple probability, we can determine the chance that a random sequence will match one from a pathogen. In reality, this approach may underestimate the true chance because all life is related and pathogenic components evolved from those not associated with pathogenicity, which is reflected in sequence similarity of conserved moieties between genes that encode proteins with unrelated functions. Conversely, a statistical approach may overestimate the chance of a match because selection will favor changes in genes that tailor their proteins to the function they are performing (that is, evolution will push for molecular divergence between two genes that encode for proteins of different functions as each gets better at its specialized job).

To determine the probability of an exact match of a sequence of any length (n) to any other sequence, the probability (p) is simply p=(0.25)n because there are four random choices for each base. To determine the probability of a sequence of any length (n) matching any other sequence with a few mismatched bases (m), the probability is:

$$P = \frac{(n!)(3^m)}{m!\,(n-m)!\,(4^m)}$$

Given that roughly 30 million CSNA orders are made each year, we can estimate how many orders will match any pathogen database if we know the content of the database. For this discussion, let us start with a very small database that contains the complete genomes of 100 viruses. Some viruses have large genomes (up to 200 kb) others are very small (less than 10 kb). For this estimate, we will assume that the "average" virus of interest to us has a 20 kb genome, and therefore the entire database would contain 2 million basepairs. This type of database may be the one chosen if de novo viral synthesis is the main risk to be addressed by the screening system. Other database options may include only the sequences of the CDC List A pathogens, which would contain approximately 12 million base pairs[152] or a more expansive database that contains 20 bacterial pathogens and 100 viral pathogens, which would house approximately 100 million base pairs. This larger database is an approximation of the data required to cover the USDA and CDC select agents which names roughly 13 toxins, 46 viruses, 21 bacteria and three eukaryotes. Note that because prokaryotic genomes are several orders of magnitude larger than viral genomes, it is the bacteria in these lists that drives the database size.

The quantity of false positives that result is not only a function of the content of the databases and the length of the query sequences, but also the degree of similarity required for a hit. Given that many pathogens exhibit a great degree of sequence divergence amongst strains of the same species, it is probably desirable for a screening tool to catch sequences from a variety of strains that could be useful to an adversary. One way to address this sequence diversity is to include sequences from all of the strains of concern (which would increase the database content). This option is further discussed in Chapter 5. Another option is to consider identical and very similar sequences a hit. Below, we present the sequence identity required to produce a hit that would lead to 1,000 serendipitous false hits a year for the three sequence databases described above. Note that as the query sequence lengthens, or the database gets smaller, the tolerance for mismatch can also drop to maintain the same false positive rate. These parameters can all be adjusted by policy-makers to tailor the final tool to the needs of the security community and to industry.

[152]Assuming one exemplar of *B. anthracis* (5.2Mb), *F. tularensis* (1.9 Mb), *Y. pestis* (4.6Mb), Ebola virus (19Kb), Marburg virus (19Kb), and smallpox virus (186Kb). Adding other hemorrhagic fever viruses would not increase the size of the database significantly.

Table 2. Percent identity required that will lead to at least 1,000 false positives a year if 25 million query sequences are searched against a database of List A agents, a database of 100 viruses, or a Select-Agent-like database.

Database	Database Size	20nt	30nt	40nt
100 Virus Database	2 million bp	95%	85%	77%
List A Database	12 million bp	100%	87%	79%
Select-Agent-like Database	100 million bp	100%	90%	80%

It is clear that, except for very close sequence matches or very long primers, many orders per year will serendipitously match sequences in a given pathogen database if anything but exact matches are required. Given the genetic polymorphisms observed even in closely related strains of bacteria, some mismatch tolerance is necessary to account for naturally occurring variants. Further complicating matters is the fact that a hostile actor could intentionally alter the wobble position of codons in a gene, potentially altering the nucleotide sequence of about every third base (perhaps reducing the nucleotide match to at best 66 percent). To counter an adversary changing the wobble position (and to reduce the impact of polymorphisms) screening systems could translate any order into six possible amino acid sequences (because from an oligo order, the start of translation will not be obvious) and check these against the sequence databases. However, because the amino acid sequences that could result from any CSNA sequence are three-fold shorter, the chances of any random sequence matching the database are greater for a sequence translated into amino acids (by a few orders of magnitude, for exact and slightly mismatched sequences of any length), even though there are 20 naturally occurring amino acids.

In reality, false hits are more likely to occur than this statistical estimate because a sequence from a non-pathogenic microbe can be very similar to one from a pathogenic one. We can expect a high-degree of similarity between the genes required for the basic metabolic functioning of the cell between pathogens and their related, non-pathogenic, counterparts. For instance, *B. anthracis* and other members of the *Bacillus* genus of soil bacteria are very closely related with a high degree of overlap in their genomes.[153,154] Read et al. indicate that there may be as few as "150 differences in the 5,000 plus genes in the chromosomes of these related bacterial species."[155] Distinguishing between the bacteria in this family has been cited as a difficulty in minimizing the false hits of screening software.[156] Parsing these differences will require fine-tuning of the search algorithms. This is further complicated by the necessity of querying many sequences against many pathogens at once, systemwide. For example, tweaking a system in order to detect differences that enable the discrimination between *B. anthracis* and *B. cereus*, may lead to too many false positives for other organisms.

One possible solution to this problem is to develop a carefully curated database which includes only the portions known to be associated with virulence from pathogen genomes in the screening database. It may also be possible to search a subsidiary database of closely related organisms that if matched would subsume any lesser scoring pathogen match, we discuss this issue further when discussing the notional system in Chapter 5.

---

[153] J. Ventner Institute, Anthrax: "A Soil Bug Gone Bad" *Scientists Decipher, Analyze Genome of Bacillus anthracis. Last accessed July 24,2007.* http://www.tigr.org/news/pr_04_30_03.shtml

[154] Ivanova, J. et al. Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis. 2003. Nature 423:87-91.

[155] Read, et al., he genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. 2003. Nature 423:81-86.

[156] Personal Communication. Robert Jones, Craic Computing.

From the previous discussion, one might wrongly conclude that false hit rates are a function of the screening tool used and, therefore, the false positive rate of the system is inextricably linked to set parameters of the mathematical algorithms that screening tools employ. We must stress that the designers of an oversight system have nearly complete control over the system false positive rate in that they can choose what degree of match is required for a hit and what the screening system should consider "risky" (just complete viral genes, just genes that are known to be associated with the ability of a pathogen to cause disease or oligos and genes that match sequences from any pathogens). To arrive at a requirement for the system false positive rate, those implementing the oversight regime will have to balance the false positive rate that would not cause undue burden to the CSNA industry and the desirable reduction of the risk that CSNA will be misused. These considerations must therefore be informed by a thorough process-based risk assessment that will help identify activities that a CSNA screening system must focus on to best reduce risk of misuse.

Conclusions

Customers and industry have clear requirements of a sequence screening tool. The tool must work within a few minutes (for the screening of oligos) to prevent the waste of materials and time for denied orders and must provide unambiguous results. Sequence screening tools must not require the order information to leave the company for screening and should provide the identity of screening hits for confirmatory screening.

Currently available software to screen CSNA orders primarily use the BLAST algorithm to inspect sequences. For instance, BlackWatch uses BLAST to compare a query sequence against a database of sequences of concern, using established parameters to minimize the false hit rate. We found in the Industry Review that companies either use BLAST, BlackWatch or a screening program of their own design. This finding indicates that gene companies are willing to employ these tools in their screening. We found that BLAST and BlackWatch perform similarly in identifying query sequences from a list of pathogens, except that BLAST has a greater tolerance for slightly mismatched sequences or sequences with non-matching nucleotides appended. This is not surprising because BlackWatch uses the BLAST algorithm but biases the program against such mismatches. We suspect queries of longer sequences would result in fewer discrepancies between the two screening systems due to the focus of BlackWatch on longer, gene-length sequences. This reflects and highlights the importance of screening parameters such as e-value, score and span, and how their settings can affect the actual screen

Also, given the abundance of researchers who work with pathogens and their activity in the molecular biology community, we can expect that any screening system would have to contend with over a thousand orders per day that legitimately contain elements associated with pathogenicity.

Chapter 4: Overview of Relevant Regulations

In recent years, as synthetic genomics techniques and technology have continued to develop, broad-based calls for regulation of the CSNA industry and other aspects of the field have begun to be heard, whether from within the synthetic biology community itself (such as the Berkeley proposals), environmental and other issue advocacy groups (such as the ETC Group), or government bodies (such as the NSABB Working Group on Synthetic Genomics). However, in assessing the validity of these individual proposals, or indeed, the necessity of regulation of any kind, by any means, it may prove fruitful to survey the existing regulatory landscape. An understanding of current regulations will serve to form a framework within which policymakers can assess proposals for further government or community-based regulation of synthetic genomics. Below, we propose to provide such a framework. We will present the content and (as far as we can ascertain it), the real-world interpretation and implementation of a number of laws and regulations of potential relevance to the CSNA industry, including the CDC/APHIS Select Agent Rules, Section 175 of Title 18 of the United States Code, EPA Regulations under the Toxic Substances Control Act (TSCA), and the NIH Guidelines for Research Involving Recombinant DNA Molecules, as well as potentially analogous regulations covering other dual-use materials.

Select Agent Rules (SAR)

Scope

The most salient feature in the regulatory landscape remains the Select Agent Rules ((42 C.F.R. Part 73, 7 C.F.R. Part 331, and 9 C.F.R. Part 121[157]), published jointly by the US Departments of Health and Human Services (HHS) and Agriculture (USDA) in the Federal Register on March 18, 2005.[158] These regulations drew upon authority established by Congress through the Antiterrorism and Effective Death Penalty Act of 1996[159], the USA PATRIOT (Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act[160] of 2001, and the Public Health Security and Bioterrorism Preparedness and Response Act of 2002[161]. The Select Agent program regulates the possession, use, and transfer of a specified list of select agents and toxins[162] HHS Select Agents/Toxins are agents that HHS considers as potentially severe risks to human health. These agents are regulated by 42 CFR 73. High Consequence Livestock Pathogens and Toxins, or USDA Select Agents/Toxins are agents that USDA considers to have the potential to pose a severe threat to animal or plant health, or to animal or plant products and are regulated under 7 CFR 331 and 9 CFR 121. Those agents that are deemed to pose a severe threat to animal health, animal products, *and* public health are referred to as HHS/USDA Overlap Select Agents and consequently appear on both the USDA and HHS Select Agent lists. The criteria for inclusion in the Select Agent lists were the following: the effect upon human/animal/plant health of exposure to the agent/toxin, the degree of contagiousness or virulence of the agent or potency of the toxin and the methods by the which it is transferred to humans/animals/plants, and finally the current availability and efficacy of drugs or vaccines to treat and prevent any illness caused by the agent or toxin.[163] The HHS Select Agent Program is administered by the Centers for Disease Control (CDC). Its counterpart at USDA is controlled by the Animal and Plant Health Inspection Service (APHIS).

---

[157]http://www.selectagents.gov/resources/42_cfr_73_final_rule.pdf
[158]These regulations represent the "final rule", superseding the "initial final rule" first promulgated in December 2002.
[159]http://thomas.loc.gov/cgi-bin/query/z?c104:S.735.ENR:
[160] http://www.selectagents.gov/resources/USApatriotAct.pdf
[161]http://www.selectagents.gov/resources/PL107-188.pdf
[162] Information is available at www.selectagents.gov/resources/salist.pdf
[163]Information is available at http://www.selectagents.gov/NSARFaq.htm#sec1q3

The Select Agent Rules apply not only to the intact and viable agents and/or toxins appearing on the CDC and USDA lists, but also potentially to "Genetic Elements, Recombinant Nucleic Acids, and Recombinant Organisms" derived from them. This formulation is subsequently defined as: "Nucleic acids that can produce infectious forms of any of the select agent viruses listed in paragraph (b) of [§ 73.3]", "Recombinant nucleic acids that encode for the functional form(s) of any of the toxins listed…if the nucleic acids: Can be expressed *in vivo* or *in vitro*, or are in a vector or recombinant host genome and can be expressed in vivo or in vitro", and "select agents and toxins…that have been genetically modified."[164]

Consequently, it would appear that the Select Agent Rules are not intended to regulate the use, possession, or transfer of genetic fragments that are unable, by themselves to produce a functional form of a listed agent or toxin. It remains unclear however (based upon the regulatory language employed), whether a set of fragments comprising a whole genome (or coding for a functional toxin) would fall under the SAR's purview. Nonetheless, it is the opinion of the NSABB Working Group on Synthetic Genomics that the SAR should be interpreted to "not apply until the functional infectious agent or toxin is generated. Thus, the language pertaining to nucleic acids and genetically modified entities", they argue, "aims to regulate the penultimate step to possessing an active and functional Select Agent" and to "avoid…the regulation of…key research reagents/products necessary for scientific advancement".[165]

The Select Agent Rules do not, however, apply to any "select agent or toxin that is in its naturally occurring environment provided [it] has not been intentionally introduced, cultivated, collected, or otherwise extracted from its natural source", to "non-viable…select agents or nonfunctional…toxins", or to certain toxins "under the control of a principal investigator, treating physician or veterinarian, or commercial manufacturer or distributor," provided that the amount possessed does not exceed certain specified levels.[166] Attenuated strains of Select Agents or Toxins may also be excluded from the provisions of the Rules, "based upon a determination that [it] does not pose a severe threat to public health and safety".

Exclusions/Exemptions

To apply for an exclusion from the SAR, the applying entity or individual is directed to submit a written request and supporting scientific information to the relevant lead agency (CDC for HHS Select Agents/Toxins, APHIS for USDA Select Agents/Toxins). Furthermore, certain categories of entities such as clinical or diagnostic laboratories and Federal law enforcement agencies are exempt from the requirements of the regulations provided that they abide by certain reporting and security requirements.[167]

Registration and Reporting

Unless qualifying for an exemption under § 73.5 or § 73.6, an individual or entity is forbidden to possess, use, or transfer any HHS Select Agent or toxin without a certificate of registration issued by the HHS Secretary (or any USDA Select Agent or toxin without one issued by the Secretary of Agriculture). To apply for registration, each entity must designate a Responsible Official (RO), who may submit the application electronically via the Select Agent Program website[168]. The Responsible Official must furthermore undergo a security risk assessment by the FBI. If the entity is seeking registration for an HHS Select Agent, it should direct its application to the CDC Division of Select Agents and Toxins. If

---

[164] 42 CFR Section 73.3

[165] "Addressing Biosecurity Concerns Related to the Synthesis of Select Agents", National Science Advisory Board for Biosecurity Working Group on Synthetic Genomics. December, 2006

[166] "100 mg of Abrin; 100 mg of Saxitoxin; 100 mg of Shiga-like ribosome inactivating proteins; or 100 mg of Tetrodotoxin".

[167] See 42 CFR § 73.4, § 73.5, § 73.6

[168] www.selectagents.gov/login.htm

the entity is seeking registration for USDA Select Agents or Toxins, it should instead apply to APHIS, while in the case that registration for an overlap agent is sought, the entity is free to choose the lead agency to which it will apply (in practice, typically choosing the CDC). A separate Certificate of Registration must be sought for each Select Agent an entity proposes to possess, use, or transfer.

Registered Entities

As of May 21, 2007, there were 406 entities (Government agencies and labs, public and private institutions of higher education, and private companies)[169] with personnel registered to work with select agents. CDC is the lead agency for 333 (83 percent) of these entities; APHIS is the lead agency for the remaining 73 (17 percent). Prior to the implementation of the final rule in March 18, 2005, 451 entities had applied (to CDC) for registration and 350 had been determined to require registration. The balance of applications (101) was not processed, primarily because CDC determined these applications to be unnecessary as they sought to register for something other than a select agent or toxin. All applicants determined to require registration were subsequently approved. The 350 CDC-registered entities (as of March 2005) fell into six groups: Academic/University: 105 (30%); Government—State/Local: 104 (~30%); Government—Federal: 61 (~17%); Commercial: 39 (~10%); Private Non-Profit/Research Institutions: 35 (~10%); Other: 6 (~2%)[170].

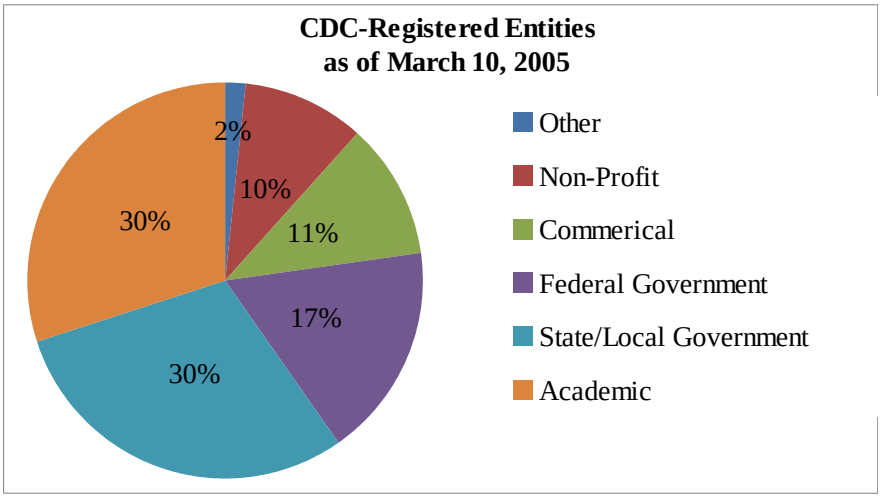


Figure 20. Classification of CDC-registered entities.

The issuance of a certificate of registration is, in practice, contingent upon the physical inspection, by agents of the CDC Division of Select Agents and Toxins, or APHIS, of the entity's facilities, security plan, biosafety plan, incident response plan, and other relevant documentation.[171] Certificates of Registration are valid for a maximum of three years.

[169] 42 CFR § 73.7. Individual laboratories within a registered entity (for example, a university), are covered by that entity's certificate of registration.

[170] "Possession, Use, and Transfer of Select Agents and Toxins; Final Rule." Title 42 Code of Federal Regulations, Vol. 70, No. 52. 2005 ed., 13315.

[171] From correspondence with Lori Bane, Compliance Officer, CDC Division of Select Agents and Toxins.

Possession, Use, and Transfer

Any registered entity is furthermore forbidden to provide any individual with access[172] to a select agent or toxin, unless that individual has been previously approved by the Select Agent Program following a security risk assessment by the FBI. Each individual with access to Select Agents or Toxins is required to have "the appropriate education, training, and/or experience to handle or use such agents or toxins"[173] and will be required to submit to a security risk assessment by the FBI.[174] Operationally, a person with "access" is defined as an individual who "get[s] his/her hands on the Select Agents or Toxins or has the ability to get his/her hands on" them.[175] Entities (excepting Federal, State, or local governmental agencies, including public accredited academic institutions), Responsible Officials, Alternate Responsible Officials, and any individuals deemed to own or control the entity are likewise subject to security risk assessments.[176] These assessments are performed at the Criminal Justice Information Service within the FBI and involve a screening system utilizing eight or nine categories of information for connections to known terrorist or criminal groups and individuals.[177] While one commenter on the interim Select Agent regulations argued that security risk assessments be expanded to include "requirements for credit checks and random drug screening", HHS ultimately concluded that Congress had not provided the Attorney General with sufficient authority to do so.[178] Instead, approval of an individual's access may be denied or revoked if he or she is found to be a "restricted person" as described by 18 U.S.C. 175b.[179] Access approval may also be denied, limited, or revoked if the individual is "reasonably suspected" by any Federal law enforcement or intelligence agency of involvement in a terrorist crime described in 18 U.S.C. 2332b(g)(5), knowing involvement with a terrorist organization (as defined in 18 U.S.C. 2331), with any other organization involved in violent crime, or of being a foreign agent (as defined in 50 U.S.C. 1801). Access may furthermore be denied, limited, or revoked if the HHS or USDA determines such action to be necessary to protect public health and safety.[180] Approval of access is valid for a maximum of five years (as compared to three years for entity registration) and is not portable should the individual transfer from one entity to another, or even to another laboratory within the same entity. However, provisions have

---

[172] "An individual will be deemed to have access at any point in time if the individual has possession of a select agent or toxin (*e.g.*, ability to carry, use, or manipulate) or the ability to gain possession of a select agent or toxin." 42 CFR § 73.10 (b)

[173] In practice, this is defined by the Select Agent Program on a "common sense" basis (from correspondence with Lori Bane). Training is defined in § 73.15.

[174] "An individual's security risk assessment may be expedited upon written request by the Responsible Official and a showing of good cause (*e.g.*, public health or agricultural emergencies, national security, or a short term visit by a prominent researcher)."

[175] From correspondence with Lori Bane, Compliance Officer, CDC Division of Select Agents and Toxins, August 6, 2007.

[176] Public accredited academic institutions are also exempt from security risk assessments of "owning" or "controlling" individuals. "For a private institution of higher education, an individual will be deemed to own or control the entity if the individual is in a managerial or executive capacity with regard to the entity's select agents or toxins or with regard to the individuals with access to the select agents or toxins possessed, used, or transferred by the entity. For entities other than institutions of higher education" and government agencies, *e.g.*, private companies, "an individual will be deemed to own or control the entity if" he or she falls into either of two categories specified in 42 CFR § 73.7 (c) (3) (ii).

[177] From phone interview with Tracy L. Rice, Management & Program Analyst, Weapons of Mass Destruction Operations Unit, WMD Directorate, FBIHQ on July 12, 2007.

[178] Federal Register, Vol. 70, No. 52, Friday, March 18, 2005. Rules and Regulations, p. 13301.

[179] "The term 'restricted person' means an individual who - (A) is under indictment for a crime punishable by imprisonment for a term exceeding 1 year; (B) has been convicted in any court of a crime punishable by imprisonment for a term exceeding 1 year; (C) is a fugitive from justice; (D) is an unlawful user of any controlled substance (as defined in section 102 of the Controlled Substances Act (21 U.S.C. 802)); (E) is an alien illegally or unlawfully in the United States; (F) has been adjudicated as a mental defective or has been committed to any mental institution; (G) is an alien (other than an alien lawfully admitted for permanent residence) who is a national of a country as to which the Secretary of State, pursuant to section 6(j) of the Export Administration Act of 1979 (50 U.S.C. App. 2405(j)), section 620A of chapter 1 of part M of the Foreign Assistance Act of 1961 (22 U.S.C. 2371), or section 40(d) of chapter 3 of the Arms Export Control Act (22 U.S.C. 2780(d)), has made a determination (that remains in effect) that such country has repeatedly provided support for acts of international terrorism; or (H) has been discharged from the Armed Services of the United States under dishonorable conditions." 18 U.S.C. 175b

[180] 42 CFR § 73.10

been made for security risk assessment portability in the case of visiting researchers, provided that both the home and host entity follow certain prescribed procedures.[181] Between April 12, 2003, when the security risk assessment process was implemented and March 10, 2005 (the date of the most recent statistics found), 8,394 staff at CDC-registered entities had received security risk assessment approval. The number of staff with approved access at registered entities was found to range from approximately five individuals at smaller facilities to one hundred or more at some large universities and commercial facilities.[182]

Registered (and all other entities) are furthermore forbidden from conducting certain "restricted" experiments with select agents or toxins unless approved by and conducted in accordance with any conditions set by the relevant lead agency. Two types of experiments are regarded as restricted: "experiments utilizing recombinant DNA that involved the transfer of a drug resistance trait to select agents that are not known to acquire the trait naturally, if such acquisition could compromise the use of the drug to control disease agents in humans, veterinary medicine, or agriculture" and those "involving the deliberate formation of recombinant DNA containing genes for the biosynthesis of select toxins lethal for vertebrates at an $LD_{50} < 100$ ng/kg body weight."[183] Any entity or individual interested in conducting experiments restricted by these regulations is required to submit a written request and supporting scientific information to the relevant lead agency. Failure to comply with the requirements determined by the relevant lead agency may lead to revocation of approval to conduct restricted experiments or to the suspension or revocation of a certificate of registration.

Transfer of select agents or toxins is only permitted between two entities registered (or exempted from registration requirements) to work with the agent or toxin. To obtain authorization for transfer, the transferring entity is required to file a completed APHIS/CDC Form 2 to the relevant lead agency. The recipient is also required to submit a completed APHIS/CDC Form 2 within two business days of receipt of a select agent or toxin. The recipient is required to immediately notify its lead agency should it fail to receive a select agent or toxin within 48 hours of the expected delivery time, or if upon receipt, the package is discovered to have been damaged to the extent that release of the select agent or toxin may have occurred. An authorization for transfer is considered valid for 30 days. With regard to international collaborations involving Select agents or toxins, the NIH "now requires that collaborative international projects meet the requirements of the select-agent regulations, including personnel screening and laboratory physical security and inspections."[184] In two types of instances, the requirements for transfer may be dispensed with: firstly, for a select agent or toxin that is contained in a specimen for proficiency testing[185], provided that at least seven days prior to the transfer, the sender reports to CDC or APHIS the select agent or toxin to be transferred and the name and address of the recipient.[186] Secondly, HHS or USDA may, on a case-by-case basis, choose to authorize a transfer under conditions other than those described above. Registered entities are moreover required under § 73.17, to maintain records pertaining to their acquisition, use, and storage of select agents and toxins.

---

[181] http://www.selectagents.gov/sra.htm

[182] "Possession, Use, and Transfer of Select Agents and Toxins; Final Rule." Title 42 Code of Federal Regulations, Vol. 70, No. 52. 2005 ed., 13315.

[183] 42 CFR § 73.13

[184] Atlas, RM, "Securing Life Sciences Research in an Age of Terrorism", *Issues in Science and Technology*. 2006.

[185] Defined as "the process of determining the competency of an individual or laboratory to perform a specified test or procedure." 42 CFR § 73.1

[186] Awaiting clarification from Lori Bane on whether the recipient must in this case, nonetheless be registered to work with the select agent/toxin.

Enforcement

Section 73.18 of the Select Agent Rules confers upon the HHS Secretary the authority to order surprise inspections of the facilities and records of any registered entities (including the ability to copy any records relating to activities covered by these regulations), to ensure compliance with the Rules. HHS is further authorized to inspect and evaluate the premises and records of any entity applying for registration, prior to the issuance of a certificate of registration. According to recent correspondence with CDC officials, the CDC alone has conducted over 630 inspections since 2003 to ensure that entities are following appropriate safety and security measures, as spelled out in the regulations. All CDC-registered entities have been inspected at least once. In the past fiscal year (2006), CDC conducted 242 inspections to register or re-register entities. Approximately 30 inspectors (both civil servants and contractors) are currently employed by the CDC Division of Select Agents and Toxins to perform these inspections.[187]

APHIS inspection procedures are similar to those of the CDC, though APHIS inspectors are not "centrally located within [the] Select Agent Program."[188] While CDC inspectors are based within the CDC Select Agent Program and write the inspection reports as well as perform the inspections, the approximately 50 APHIS inspectors complete standard checklists and submit them to the APHIS Select Agent Program for review. The contents of these checklists are then incorporated into inspection reports generated within the APHIS Select Agent Program.

In the event that suspicion of a safety or security violation arises, whether in course of inspections, or by some other means (such as a tip), the regulatory authorities are able to apply administrative, civil, and/or criminal penalties upon violators. Administrative penalties (denial, suspension, or revocation of registration) are dispensed by CDC or APHIS and can be appealed in writing within 30 calendar days of the decision. Upon revocation or suspension of a certificate of registration, the entity is required to "immediately stop all use of each select agent or toxin covered by the…order…safeguard and secure each select agent or toxin…from theft, loss, or release, and comply with all disposition instructions issued" by the relevant lead agency.[189] Regarding civil penalties, the Office of the Inspector General of the Department within HHS (and its analogue within USDA) is delegated authority to conduct investigations and to impose civil money penalties against any individual or entity for violations of the Select Agent Rules, as authorized by the Public Health Security and Bioterrorism Preparedness and Response Act of 2002. Penalized individuals and entities are entitled to appeal to the relevant Departmental Appeals Boards of the Departments of Health and Human Services and Agriculture, who are empowered to make final determinations with regard to civil money penalties for violations of these regulations. Civil penalties may be pursued in conjunction with criminal ones.

In the event that a suspicion of criminal misconduct arises on the part of CDC/APHIS or the Inspector s General of HHS and USDA, the case will be referred to relevant officials with the FBI.[190] Criminal penalties for violations of the Select Agent regulations are defined under Title 18 Section 175b of the US Code. Possession without registration of a select agent or toxin is punishable by up to five years' imprisonment and/or a fine. Transfer to unregistered persons (if the transferor knows, or has reasonable cause to believe recipient is unregistered) is punishable by up to five years' imprisonment and/or fine. Furthermore, the transport or shipment (excluding "duly authorized" US Government activity) of select

[187] From Correspondence with Lori Bane, Compliance Officer, CDC Division of Select Agents and Toxins, various dates.
[188] From Correspondence with Lori Bane, Compliance Officer, CDC Division of Select Agents and Toxins, August 6, 2007.
[189] 42 CFR § 73.8
[190] Should suspicion of criminal activity arise, CDC/APHIS would be expected to immediately suspend registration and to contact the FBI through a well-defined and robust liaison system. However, to the best knowledge of our contacts at the FBI WMD Directorate and the CDC Select Agent Division, suspicion of criminal misconduct has yet to arise in the course of CDC or APHIS investigations of Select Agent Rule violations. (From interviews and correspondence with Lori Bane of CDC and Tracy Rice of FBI).

agents or toxins via interstate or international commerce (or the receipt of select agents or toxins through interstate or international commerce) is punishable by up to 10 years' imprisonment and/or fine.[191] Section 175 of Title 18 of the US Code prescribes criminal penalties of fines and/or imprisonment for up to 10 years, for the possession of "any biological agent, toxin, or delivery system of a type or in a quantity that, under the circumstances, is not reasonably justified by a prophylactic, protective, bona fide research, or other peaceful purpose".[192] Possession of such materials for "use as a weapon" is, for its part, punishable by a fine, life imprisonment, or both.[193]

Production, engineering, synthesis, or acquisition of variola virus (defined as "a virus that can cause human smallpox or any derivative of the variola major virus that contains more than 85 percent of the gene sequence of the variola major virus or the variola minor virus")[194] are prohibited under Section 175c. Depending upon the severity of the offense, violations are punishable by fines up to $2 million and sentences ranging from 25 years-to-life to the death penalty. It is worth noting that *any* work with variola virus (as defined in the statute), outside that "conduct[ed] by, or under the authority of the Secretary of Health and Human Services"[195] is expressly forbidden, regardless of its purpose. This may prove problematic as there are many regions within the variola major and variola minor genomes with significantly greater than 85 percent homology with related, but relatively harmless viruses such as vaccinia that are vital to beneficial research. Indeed, a strict interpretation of the current statute could inadvertently criminalize many types of vital and constructive research, up to, and including the development and production of a smallpox vaccine. As *Science* magazine noted in March of 2005, "virologists zooming in on the [law's] fine print…cannot agree on what exactly it outlaws."[196] Peter Jahrling, a researcher "who worked with variola at the US Army Medical Research Institute of Infectious Diseases [USAMRIID] in Fort Detrick, Maryland" opined that "an overzealous interpretation 'would put a lot of poxvirologists in jail'."[197] On the other hand, UNMOVIC Bioweapons experts Roman Mezencev and Kay Mereish argued in a subsequent letter to *Science* that the proper usage of alignment techniques[198] could prevent beneficial research from being wrongly categorized as illegal.[199] Nonetheless, it was with concerns of excessive restrictions on legitimate research in mind that the NSABB Working Group on Synthetic Genomics recommended the repeal of 18 U.S.C. § 175c, arguing that its sequence homology stipulation was arbitrary, unduly restrictive to beneficial research and moreover, unnecessary, as the misuse of variola remains covered under other subsections of § 175. The Working Group's recommendations were subsequently unanimously approved by the NSABB, to the consternation of advocacy groups such as The Sunshine Project and the ETC Group, who decried the decision as "shocking", "dangerous", and "unfortunate".[200]

Procedures for government seizures of potentially dangerous biologicals are described under Section 176 of Title 18 of the U.S. Code. Under this section, the Attorney General is empowered to request a warrant for the seizure of any biological agent, toxin, or delivery system that "pertains to conduct prohibited under

---

[191] 18 U.S.C. § 175b
[192] 18 U.S.C. § 175 (b)
[193] 18 U.S.C. § 175 (c)
[194] 18 U.S.C. § 175c (d)
[195] 18 U.S.C. § 175c (a) (2)
[196] Enserink, M. "Unnoticed Amendment Bans Synthesis of Smallpox Virus" *Science*, **307** (5715), 2005.
[197] Enserink, M, 2005.
[198] Mezencev and Mereish point out that while camelpox virus strain CMS and vaccinia virus strain Copenhagen "were shown to share nucleotide identity [of] 91%" with "variola major virus strain Bangladesh-1975…throughout the conserved central region of their genomes", the usage of a global alignment of whole genomes would be more appropriate "for the interpretation of the Act…than the local alignment of their conserved portions". They cite data from "global alignment by Lalign program" showing that "variola major virus (NC_001611 shares identity of only 28.2% with variola minor Garcia-1966, 28% with vaccinia virus WR, and 27.9% with Camelpox virus CMS."
[199] Mezecev, R and Mereish, K, "How Similar are Poxviruses?" *Science*, **308** (5726), 2005.
[200] ETC Group, "Extreme Genetic Engineering: An Introduction to Synthetic Biology" p.159, January 2007.

section 175" of the title or that "has no apparent justification" for some peaceful purpose. The Justice Department is also authorized, "in exigent circumstances" to order the seizure of such materials without a warrant.[201] Notably, the term "biological agent" is subsequently defined (in § 178 of the title) as "any microorganism (including, but not limited to, bacteria, viruses, fungi, rickettsiae or protozoa), or infectious substance, or any naturally occurring, bioengineered or *synthesized component of any such microorganism or infectious substance*, capable of causing death, disease, or other biological malfunction in a human, an animal, a plant, or another living organism; deterioration of food, water, equipment, supplies, or material of any kind; or deleterious alteration of the environment."[202,203] Likewise, "toxin" is defined such that it encompasses any toxic products of a "recombinant or synthesized molecule", whatever their origin and method of production.

Violations of the Select Agent Rules may also be prosecuted, where relevant, under other federal statutes governing terrorism (18 U.S.C. § 2331), weapons of mass destruction (18 U.S.C. § 2332a), conspiracy (18 U.S.C. § 371) mail fraud (18 U.S.C. § 1341), wire fraud (18 U.S.C. § 1343), and the use of a false name and address (18 U.S.C. § 1342), among others. In the hypothetical instance that an individual misrepresented his or her identity or intentions to obtain select agents or toxins, he or she might be subject to penalties of fines *or* imprisonment of not more than 20 years (under 18 U.S.C. § 1341 and/or 18 U.S.C. § 1343) for mail or wire fraud and additional fines *or* imprisonment of up to five years (under 18 U.S.C. § 1342) for the use of a false name and address. Naturally, in the event that a select agent or toxin is used or stockpiled for use as a weapon of "mass destruction" or "terror", whether resulting in death or not, the prescribed penalties for the perpetrator become correspondingly harsher, escalating to life imprisonment and the death penalty.

Costs

The Select Agent Rule does not trigger the "unfunded mandate" executive order (12866). CDC estimates the total annualized cost of the final rule at $16 million, with annualized costs per facility ranging from $15,300 to $170,000. CDC's analysis of the Select Agent Rule's costs and benefits concluded that the regulations do not create an unfunded mandate, noting that the select agent program imposes "biosecurity and physical security requirements…that [CDC] believe[d] were already industry standards", such as the biosecurity standards recommended by the Biosafety in Microbiological and Biomedical Laboratories, 4th Edition (BMBL), published by the CDC and NIH.[204] Congress authorized "such sums as may be necessary" to create, implement, and enforce the select agent regulations under Section 351A (m) (Enhanced Control of Dangerous Biological Agents and Toxins) of the "Public Health Security and Bioterrorism Preparedness and Response Act of 2002".

---

[201] 18 U.S.C. § 176

[202] 18 U.S.C. § 178

[203] The precise meaning of this statutory language is ambiguous to the layperson. It could potentially cover only those "synthesized component of any…microorganism or infectious substance" that are able to cause "death, disease, or other biological malfunction" etc., *on its own*, or merely any synthesized component of a whole microorganism or infectious substance that was capable of causing the described effects.

[204] "Possession, Use, and Transfer of Select Agents and Toxins; Final Rule." Title 42 Code of Federal Regulations, Vol. 70, No. 52. 2005 ed., 13316.

EPA Regulations

Scope

The EPA regulates the development and production of "new"[205] microbes "for commercial purposes" created via trans-generic[206] recombinant genetics under authority from the Toxic Substances Control Act (15 U.S.C. 2615). Anyone intending to "manufacture, import, or process" microorganisms for commercial purposes is required to file either a "Microbial Commercial Activity Notice (MCAN)" with the EPA, or if the microorganism is intended for use in research and development activities, a TSCA Experimental Release Application (TERA) for a specific test (involving release of the microorganism into the environment).[207] "Microorganism" is defined to mean "an organism classified, using the 5-kingdom classification system of Whittacker, in the kingdoms Monera (or Procaryotae), Protista, Fungi, and the Chlorophyta and the Rhodophyta of the Plantae, and a virus or virus-like particle."[208] "Commercial Purposes" are defined to include any activities "with the purpose of obtaining an immediate or eventual commercial advantage for the manufacturer, importer, or processor" and covers the usage of "any amount of a microorganism or microbial mixture". Commercial distribution, including test marketing, product research, and the development of an intermediate are also covered under these regulations. Notably, the "term also applies to substances that are produced coincidentally during the manufacture, processing, use, or disposal of another microorganism or microbial mixture, including byproducts that are separated from [it]…and impurities that remain in [it]." Perhaps of particular interest is the stipulation that "mobile genetic elements", defined as any "element of genetic material that has the ability to move genetic material within and between organisms…. Includ[ing] all plasmids, viruses, transposons, insertion sequences, and other classes of elements with these general properties" be regulated as well.[209] Needless to say, the breadth of these regulations is such that essentially any conceivable commercial activity (involving any type of recombinant trans-generic microorganism) undertaken by the biotechnology industry would necessarily fall under their purview. However, in the case of certain research and development activities, these regulations are entirely superseded by other federal rules and regulations. In many cases, work with Select Agents, for example, may not be subject to EPA regulation at all, provided that it meets *all* of a specified number of criteria.[210]

Enforcement

The EPA conducts both regularly scheduled and surprise inspections of MCAN and TERA-filing companies. According to recent conversations with EPA officials, the agency currently employs approximately 30 inspectors each in 4 of the EPA TSCA Biotechnology program's 10 regional offices (2, 4, 5, and 8).[211, 212] Every MCAN or TERA-filing company has been inspected at least once. Inspections

---

[205] Defined as "those microorganisms formed by deliberate combinations of genetic material from organisms classified in different taxonomic genera" at http://usbiotechreg.nbii.gov/roles.asp.

[206] Interview with Thomas Crosetto of EPA TSCA Region 5 (IL,IN,MI,MN,OH,WI), July 26, 2007.

[207] 40 CFR § 725.1

[208] 40 CFR § 725.3

[209] 40 CFR § 725.3

[210] "The microorganism is manufactured, imported, or processed solely for research and development activities. There is no intentional testing of a microorganism outside of a structure, as a structure is defined in § 725.3. The person receives research funds from another Federal agency, and the funds are awarded on the condition that the research will be conducted in accordance with the relevant portions of the NIH Guidelines, or a Federal agency or program otherwise imposes the legally binding requirement that the research is to be conducted in accordance with relevant portions of the NIH Guidelines." 40 CFR § 725.232

[211] Interview with Michael Bias, CBI Coordinator, EPA TSCA Region 2 (NJ, NY, PR, VI) on July 19, 2007.

[212] EPA TSCA Biotechnology Program Contacts (very out of date) are available here: http://www.epa.gov/opptintr/biotech/pubs/biocontx.htm

however, are restricted by statute (15 U.S.C. § 2610) from extending to "financial data, sales data (other than shipment data), pricing data, personnel data, or research data" (other than data required to verify compliance). In the event that violations are discovered in the course of an inspection, companies may be subject to civil penalties of up to $25,000 for each violation of the regulations.[213]

Because of the breadth of activities that this regulation appears to cover, one would expect that almost any company (except those that receive government funding) using microbes in the course of their research and development activities, even if they never intend to sell those microbes, would fall under this regulation. However, given that very few companies are regulated under, or even aware of, this regulation, we assume that enforcement of the regulation focuses on only companies that are developing a microbe for release into the environment. However, despite repeated requests for clarification, we have not been able to verify this interpretation with the EPA.

Registration and Reporting

The provisions for reporting and the protection of "confidential business information", or CBI are likely to prove the most relevant to our concerns. Entities making MCAN or TERA submissions to the EPA under these regulations are required to submit information allowing the microorganism to be "accurately and unambiguously identified", including taxonomic designations "for the donor organism and the recipient microorganism to the level of strain, as appropriate. These designations must be substantiated by a letter from a culture collection, literature references, or the results of tests conducted for the purpose of taxonomic classification." The submitting entity is furthermore required to provide, upon the EPA's request, data supporting the taxonomic designation, including "the genetic history of the recipient microorganism…documented back to the isolate from which it was derived."[214] Submitters are moreover directed to provide "supplemental" information incorporating both phenotypic[215] and genotypic[216] information.[217] As the preceding potentially encompasses a great deal of sensitive information that submitters may wish to keep confidential, they are permitted to assert a claim of confidentiality if they believe "that public disclosure prior to commencement of manufacture or import for general commercial use, or the fact that anyone is initiating research and development activities pertaining to the specific microorganism or intends to manufacture or import the specific microorganism for general commercial use would reveal confidential business information".[218] Such an assertion must be substantiated according to the requirements given in 40 CFR § 725.94. In the case that such an assertion is accepted, CBI pertaining to the submission will not be made public and will instead be replaced by generic information. As approximately 95 percent of MCAN and TERA submission are made on paper (the remaining 5 percent are submitted as electronic documents contained on compact discs), the security of any CBI contained therein (or in supporting documents) is guaranteed by physical means. No one is permitted to hold CBI overnight and it must be logged in and out of the secured safes in which it is stored. Complete annual inventories are made of all CBI stored in EPA archives.[219] While the relevance of these practices for the proposed storage of sequence screening data is somewhat limited by the fact that CBI stored by the EPA exists almost completely on paper (thus obviating the need for secure electronic data

[213] 15 U.S.C. § 2615
[214] 40 CFR § 725.12 (a)
[215] Phenotypic information is defined as "pertinent traits that result from the interaction of a microorganism's genotype and the environment in which it is intended to be used and may include intentionally added biochemical and physiological traits". 40 CFR § 725.12 (b) (1)
[216] Genotypic information is defined as "the pertinent and distinguishing genotypic characteristics of a microorganism, such as the identity of the introduced genetic material and the methods used to construct the reported microorganism. This also may include information on the vector construct, the cellular location, and the number of copies of the introduced genetic material." 40 CFR § 725.12 (b) (2)
[217] 40 CFR § 725.12 (b)
[218] 40 CFR § 725.85
[219] Interview with Michael Bias, CBI Coordinator, EPA TSCA Region 2 (NJ, NY, PR, VI) on July 19, 2007.

transmission and storage), it is nonetheless important to note that the Federal Government does currently employ a working system for the storage of CBI related to potentially hazardous biological products.

Costs

EPA Regulations under TSCA: Regulatory Impact Analysis prepared by EPA prior to the issuance of the final rule on April 11, 1997 found that the regulations did not qualify as an "unfunded mandate" under executive order 12866. EPA estimated at the time that approximately 130 firms and as many as 300 universities could be affected by the regulations. However, the number of affected entities in both industry and academe are likely to have increased dramatically since. Annual costs to industry were estimated (in 1995 dollars) to amount to between $1.2 million and $3.0 million in the first year of implementation of the regulations and between $95,000 and $690,000 in year five.[220, 221]

Department of Commerce Regulations

Scope

There are two primary government licensing agencies in the United States with authority over the export of biological materials: the Department of Commerce, Bureau of Industry and Security, which has licensing authority over dual-use items, and the Department of State, Directorate of Defense Trade Controls, which possesses sole licensing authority over items that are considered munitions items under the International Traffic in Arms Regulations.

Dual-use biologicals controlled by Commerce under the Export Administration Act are listed under Export Control Classification Numbers (ECCNs) 1C351 ("Human and zoonotic pathogens and toxins"), 1C352 ("Animal pathogens"), 1C353 ("Genetic elements and genetically-modified organisms"), 1C354 ("Plant pathogens") and, 1C360 ("Select agents not controlled under ECCN 1C351, 1C352, or 1C354) found in Supplement No. 1 to Part 774 of the Export Administration Regulations (EAR). These regulations are based on the list developed by the Australia Group and are enforced (under the same classification system) in each member nation. Consequently, the list of agents controlled under the EAR, while substantially similar to those found on the CDC/APHIS lists, is far from identical. *Chlamydia psittaci*, for example, is controlled under the former, but not the latter. There are, moreover, substantial differences in how each set of regulations might apply to synthetic DNA.

Of particular interest in this respect is ECCN 1C353 ("Genetic elements and genetically-modified organisms), under which the export of "genetic elements", "that contain nucleic acid sequences *associated with the pathogenicity* of microorganisms controlled by 1C351 a. to c., 1C352, 1C254, or 1C360[222]" as well as those that "contain nucleic acids coding for any of the 'toxins' controlled by 1C351.d or any 'sub-units of toxins' thereof" is restricted. "Genetic elements" are moreover, defined to "include, *inter alia*, chromosomes, genomes, plasmids, transposons, and vectors, whether genetically modified or unmodified". "Genetically modified organisms", for their part, are defined thusly: "organisms that contain nucleic acid sequences associated with the pathogenicity of microorganisms

---

[220] "Microbial Products of Biotechnology; Final Regulation Under the Toxic Substances Control Act; Final Rule." Title 40 Code of Federal Regulations, Vol. 62, No. 70. 1997 ed., 17929.

[221] Year Five costs account for costs associated with rule familiarization only in the case of new firms entering the affected market areas, and are therefore much lower than year 1 costs, which account for all existing affected entities.

[222] This particular formulation is defined to mean "any sequence specific to the relevant controlled microorganism that: a. In itself or through its transcribed or translated products represents a significant hazard to human, animal or plant health; or b. Is known to enhance the ability of a microorganism controlled by 1C351 a. to c., 1C352, 1C354, or 1C360, or any other organism into which it may be inserted or otherwise integrated, to cause serious harm to human, animal or plant health." ECCN 1C353, Supplement No. 1 to Part 774 of the Export Administration Regulations.

controlled by 1C351 a. to c., 1C352, 1C254, or 1C360" or "coding for any of the 'toxins' controlled by 1C351.d or 'sub-units of toxins' thereof." Of additional interest related to this point is the potential for international acceptance of guidelines relating to tightened control of GMOs. The EU has traditionally held a stronger stance on GMOs than the United States, and if synthetic biology oversight guidelines can be applied to GMOs as well, then that would likely bolster their appeal and speed their adoption by the EU.

It is worth noting that unlike many other portions of the Export Administration Regulations, the ECCNs governing the export of dangerous human, animal, and plant pathogens and toxins (along with the genetic material associated with them) remain in force regardless of the intended destination of the export. In other words, exports to Canada, the EU, or Japan require that the exporter undergo the same licensing as if it intended to export to China or Saudi Arabia.

Furthermore, even if all exports of dual-use biologicals (such as genetic material "associated with pathogenicity" from ECCN-listed agents) were to immediately cease, the Department of Commerce would likely continue to exercise regulatory authority over the synthetic oligo industry through its control over "deemed exports" of technology (broadly defined in this context to mean "specific information necessary for the 'development,' 'production,' or 'use' of a product" controlled under the EAR)[223] via the transfer of expertise to a non-resident foreign national (including those in the United States under H-1B work, not student, visas). As in the case of "physical" exports, the "exporter" is required to seek an export license. "Exporters" are exempted from this requirement if the technology transfer occurs in the course of "fundamental research", which is pursued without a specific practical aim or objective with the intention of publication in the scientific or academic literature.[224] However, research and development conducted by private corporations, or funded by corporations, in which the findings are reviewed with the intent of controlling the results to be released in the open literature are considered proprietary and are subject to the licensing requirement.[225]

It is important to note that Commerce Regulations covering biologicals (Category 1C) have no "Low Value Shipment" authorized.[226] That is, a shipment of pathogens or genetic material will fall under the regulations regardless of its value. For other controlled items, like manufacturing equipment, the item must be worth at least a certain value for Commerce Department Regulations to apply.

Enforcement

The issuance of an export license from either agency requires that all relevant information on the item to be exported and the end user have been submitted by the exporter and the completion of a thorough government review of proposed export. In practice, the licensing process typically takes from six to eight weeks in the case of Commerce-regulated (dual-use) exports, while munitions exports controlled by the State Department may require a period of several months to gain approval.[227] Unless a license has been approved, the shipment of an item without said license is illegal. If we are interpreting this rule correctly and sequences derived from pathogens are as prevalent as we believe, the larger CSNA suppliers probably violate this rule on a daily basis. Customs and Border Protection has primary responsibility for the investigation of export violations and will likely be the first to investigate when an item is found to lack the necessary paperwork. In addition the Office of Export Enforcement within the Department of

[223] http://www.bis.doc.gov/DeemedExports/DeemedExportsFAQs.html#23
[224] EAR Part 772, page 4
[225] http://www.bis.doc.gov/DeemedExports/DeemedExportsFAQs.html#23
[226] Joe Chuchla, former Director, Nuclear Technology Division, Bureau of Export Administration (now Bureau of Industry and Security), Department of Commerce, personal communication, September 19, 2007.
[227] From conversations and correspondence with Joe Chuchla, former Director, Nuclear Technology Division, Bureau of Export Administration (now Bureau of Industry and Security), Department of Commerce, May 25, 2007 – June 11, 2007.

Commerce, Bureau of Industry and Security may also investigate and file charges for an illegal export. In either case, criminal and or civil charges may be filed and could involve large fines, prison time and/or the denial of all export privileges.

Under the present guidelines, all export license applications that are submitted to the US Department of Commerce, Bureau of Industry and Security (BIS) are reviewed by the departments of State and Defense and in some cases Energy and Justice. These applications may also reviewed by the Non-Proliferation Center (NPC) of CIA. License reviews conducted by the Departments of State and Defense may also include an intelligence review, typically performed by the Defense Intelligence Agency (DIA). A further internal review of the export license application is performed within BIS by the Office of Export Enforcement which has an independent review of the application and can put a hold on the application should it uncover any areas of potential concern.

The review process typically must be completed within 30 days of the agencies' receipt of the export application for review. This timeframe, combined with the technical difficulties discussed above make this rule not well suited for the CSNA industry.

Due to the number of export applications received by BIS (about 12,000 yearly), NPC has restricted their review to a few areas of concern such as exports to the People's Republic of China, etc. It is consequently difficult to imagine this review program being substantially extended to cover exports of dual-use biologicals such as synthetic nucleotides, total orders for which number in the tens of millions each year, even if the industry first determines which few percent of CSNA products exported match sequences from controlled pathogens.

Furthermore, of the companies we talked to, none claimed to have obtained an export license for the export of CSNA. Clearly, some of the millions of CSNA orders shipped overseas would be "associated with pathogenicity" under the strictest definition of pathogenicity, and should require an export license, no matter which country they are being exported to. The fact that no company seems to be obtaining an export license suggests that these regulations are not enforced for CSNA. Even if CSNA products were to be regulated, given the difficulty, without destructive testing, of identifying the dual-use nature of CSNA, it would be nearly impossible for customs inspectors to identify an illicit package unless the shipper declares the contents accurately.

At the present time, the law governing dual-use exports under Commerce control has expired and the regulations are being governed by an emergency powers Act. As seen in Figure 21 below, there is proposed language to enact a new Export Administration Act with higher fines and penalties.

Existing EAA	**IEEPA**	**Export Enforcement Act of 2007**
Criminal Maximum Corporate. Greater of $1 million or five times the value of exports involved Individual. $250,000 and 10 years imprisonment	Criminal Maximum Corporate. $50,000 (greater or twice the gross gain or loss, or $500,000 under alternative criminal code fine provision) Individual. $50,000 and 10 years imprisonment (greater of twice the gross gain or loss, or $250,000 under alternative criminal code fine provision)	Criminal Maximum Corporate. Greater of $5 million or ten times the value of the exports involved Individual. Greater of $1 million and 10 years imprisonment
Civil Maximum $11,000 per violation ($120,000 for national security violations)	Civil Maximum $50,000 per violation	Civil Maximum $500,000 per violation

Figure 21. Existing, proposed and emergency penalties for the violation of Export Administration Regulations.

NIH Guidelines

While they are not always invested with binding regulatory authority, the NIH Guidelines for Research Involving Recombinant DNA Molecules (NIH Guidelines), are perhaps the most relevant feature of the regulatory landscape, as they are the regulations with which the vast majority of researchers are most familiar. The NIH Guidelines are intended to specify safety practices and containment procedures for research involving recombinant DNA,[228] including the creation and use of organisms and viruses containing recombinant DNA. Institutions involved in conducting or sponsoring any recombinant DNA research funded in part, or whole, by the NIH are required to adhere to the Guidelines.[229] In addition, funding from other federal agencies or private sources may often be contingent on compliance with the Guidelines. Indeed, the NIH Guidelines are widely regarded as de facto standards within the research community and are often implemented by researchers (such as those with wholly private funding or those based in other countries) who would not otherwise be obligated to do so.

Under the Guidelines, experiments involving recombinant DNA are classified into six categories,[230] based upon the number of regulatory hurdles they are required to clear to receive approval. For example, experiments involving the deliberate transfer of a drug resistance trait to microorganisms not known to acquire the trait naturally, resulting in the potential compromise of efficacy for drugs used to control disease agents in humans, animals, or plants are to be considered "Major Actions" under the Guidelines, requiring Institutional Biosafety Committee [IBC] approval, NIH Recombinant DNA Activity Committee review, and NIH Director approval before initiation.[231] Experiments involving the cloning of Toxin Molecules with $LD_{50} < 100$ ng/kg of vertebrate body weight are regulated under Section III-B-1 of the Guidelines and are deemed to require the prior approval of the researcher's IBC as well as the NIH Office of Biotechnology Activities (OBA) and are to be conducted under containment conditions set by the NIH OBA in consultation with ad hoc experts. Experiments proposing the transfer of recombinant genetic material to one or more Human Research Participants are regulated under Section III-C and require IBC and IRB approval and RAC review before participants may be enrolled. Experiments involving the introduction of recombinant DNA into Agents classified as NIH Risk Group 2, 3, 4, or "Restricted" (Select) agents are determined to require IBC approval prior to initiation (though higher-risk, *i.e.*, higher-number agents require correspondingly stricter containment requirements).[232] Experiments involving the reverse process, *i.e.*, the introduction of DNA from pathogens to non-pathogenic microorganisms, are regulated under Section III-D-2 and are subject to similar requirements. However, experiments proposing to introduce recombinant DNA into Select agents, or the transfer of Select agent DNA to non-pathogens, are specifically required under Sections III-D-1-d and III-D-2-b of the Guidelines to be conducted under containment conditions determined by NIH/OBA review.

---

[228] Defined as either "molecules that are constructed outside living cells by joining natural or synthetic DNA molecules that can replicate in a living cell, or…molecules that result from the replication of those described above." "NIH Guidelines for Research Involving Recombinant DNA Molecules", April 2002, Section I-B. "Definition of Recombinant DNA Molecules".

[229] "NIH Guidelines for Research Involving Recombinant DNA Molecules", April 2002, Section I-C. "General Applicability".

[230] "(i) those that require Institutional Biosafety Committee (IBC) approval, RAC [NIH Recombinant DNA Advisory Committee] review, and NIH Director approval before initiation (see Section III-A), (ii) those that require NIH/OBA [NIH Office of Biotechnology Activities] and [IBC] approval before initiation (see Section III-B), (iii) those that require [IBC] and Institutional Review Board approvals and RAC review before research participant enrollment (see Section III-C), (iv) those that require [IBC] approval before initiation (see Section III-D), (v) those that require [IBC] notification simultaneous with initiation (see Section III-E), and (vi) those that are exempt from the NIH Guidelines (see Section III-F)." "NIH Guidelines for Research Involving Recombinant DNA Molecules", April 2002, Section III.

[231] *Ibid*., Section III-A-1-a.

[232] The Guidelines note that IBCs will typically require Biosafety Level (BSL) 2 containment for experiments utilizing Risk Group 2 agents and BSL 3 containment for experiments with Risk Group 3 agents. However, the Guidelines require that all experiments involving Risk Group 4 Agents be conducted under BSL 4 containment. *Ibid*., Section III-D-1-a – c.

Table 3. Required actions under NIH Guidelines by experiment type.		
Experiment Type	**Classification under NIH Guidelines**	**Required Actions**
Experiments Involving the Deliberate Transfer of a drug resistance trait (that could compromise the use of the drug to control disease agents in humans, animals, or plants) to microorganisms that are not known to acquire the trait naturally	Section III-A-1-a - "Major Actions"	IBC Approval, RAC Review, and NIH Director Approval *prior to* initiation of the proposed experiment
Experiments Involving the Cloning of Toxin Molecules with LD50 < 100 nanograms/kilogram body weight	Section III-B-1	NIH/OBA and IBC Approval prior to initiation of the proposed experiment
Experiments Involving the Deliberate Transfer of Recombinant DNA, or DNA or RNA Derived from Recombinant DNA into One of More Human Research Participants	Section III-C-1	IBC and IRB Approval and RAC Review *prior to* Research Participant Enrollment
Experiments Using Risk Group 2, 3, 4 or Restricted (Select) Agents as Host-Vector Systems	Section III-D-1	IBC Approval *prior* to initiation of the proposed experiment
Experiments in Which DNA from Risk Group 2, 3, 4, or Restricted (Select) Agents is Cloned into Nonpathogenic Microbes	Section III D-2	IBC Approval *prior* to initiation of the proposed experiment
Experiments Involving the Use of Infectious DNA or RNA Viruses or Defective DNA or RNA Viruses in the Presence of Helper Virus in Tissue Culture Systems	Section III-D-3	IBC Approval *prior* to initiation of the proposed experiment
Experiments Involving Whole Animals	Section III-D-4	IBC Approval *prior* to initiation of the proposed experiment
Experiments Involving Whole Plants	Section III-D-5	IBC Approval *prior* to initiation of the proposed experiment
Experiments Involving More than 10 Liters of Culture	Section III-D-6	IBC Approval *prior* to initiation of the proposed experiment
Experiments Involving the Formation of Recombinant DNA Molecules Containing ≤ Two-Thirds of the Genome of any Eukaryotic Virus	Section III-E-1	IBC Notice *simultaneous* with initiation of the proposed experiment
Experiments Involving Whole Plants	Section III-E-2	IBC Notice *simultaneous* with initiation of the proposed experiment
Experiments Involving Transgenic Rodents	Section III-E-3	IBC Notice *simultaneous* with initiation of the proposed experiment

Table 3. Required actions under NIH Guidelines by experiment type.		
Experiment Type	**Classification under NIH Guidelines**	**Required Actions**
Exempt Experiments	Section III-F-1, Section III-F-2, Section III-F-3, Section III-F-4, Section III-F-5, Section III-F-6	N/A

Applicability of Current Regulations

The way in which synthetic or recombinant DNA is defined by the various Federal regulatory authorities is of primary importance in assessing the applicability of current regulations to the synthetic oligo industry community. "Genetic Elements, Recombinant Nucleic Acids, and Recombinant Organisms" regulated under the Select Agent Rules are defined to mean "nucleic acids that can produce infectious forms of any of the select agent viruses listed in paragraph (b) of [§ 73.3]", "Recombinant nucleic acids that encode for the functional form(s) of any of the toxins listed…if the nucleic acids: Can be expressed *in vivo* or *in vitro*, or are in a vector or recombinant host genome and can be expressed *in vivo* or *in vitro*", and "select agents and toxins…that have been genetically modified."[233] As the NSABB Working Group on Synthetic Genomics notes, interpretation of this clause of the regulations requires one to parse what is meant by "can produce" as well as what constitutes a "select agent virus."[234] One might interpret the latter to mean an intact genome, but one could also conceivably define "virus" in the sense of 18 U.S.C. § 175c, which considers "variola virus" to mean not only a "virus that can cause human smallpox", but also, "any derivative of the variola major virus that contains more than 85 percent of the gene sequence of the variola major virus or the variola minor virus."[235] Moreover, the statutory definition of variola virus would seem to run counter to a straightforward reading of "Recombinant nucleic acids," which might otherwise indicate that only DNA in an expression vector could be regulated. There is further ambiguity regarding whether DNA encoding for toxins produced by Select Agents, which do not however, appear on the Select Toxins list, *e.g.*, anthrax lethal toxin should be subject to these regulations.[236] Importantly, practitioners in the field, such as the biosafety officer of a major US research institution and cleared SAR facility, interprets the Select Agent Rules as regulating the possession of DNA only when the DNA encodes an entire pathogen genome or a toxin on the list.[237]

Further clouding the already murky regulatory waters are the substantial differences between the Export Administration Regulations (Commerce Control List) and the Select Agent Rules in their descriptions of the genetic material subject to regulation. Indeed, these differences are extensive enough that one set of regulations might appear to allow the transfer of certain genetic material restricted by the other.[238] In the case of the former, restricted genetic elements are defined to include, inter alia, chromosomes, genomes,

---

[233] 42 CFR Section 73.3

[234] "Addressing Biosecurity Concerns Related to the Synthesis of Select Agents", National Science Advisory Board for Biosecurity, December 2006., p. 7

[235] 18 U.S.C. § 175c (d)

[236] It is, however, the understanding of the NSABB Working Group that the Select Agent "Rules intentionally do not apply until the functional infectious agent or toxin is generated. Thus, the language pertaining to nucleic acids and genetically modified entities aims to regulate the penultimate step to possessing an active and functional Select Agent.", p. 9

[237] Interviewee and institution to remain anonymous, interview conducted on June 20, 2007.

[238] Though the Export Administration Regulations do not cover the domestic transfer of controlled agents or toxins and thus, are not in legal conflict with the Select Agent Rules, an inconsistency in US biosecurity policy may easily prove nearly as costly as one in law.

plasmids, transposons and vectors, whether genetically modified or unmodified, and genetically modified organisms,[239] while as previously discussed, the Select Agent Rules might easily be interpreted less restrictively. To take another example, while the language of the Select Agent Rule seems to suggest, as previously discussed, that genetic material coding for a toxin produced by a listed agent that was not listed in its own right (as a Select Toxin), would not be subject to its restrictions, such material would very likely be controlled under ECCN 1C352, which regulates "any sequence specific to the relevant controlled microorganism that: in itself or through its transcribed or translated products represents a significant hazard to human, animal, or plant health." Given that, in the words of the NSABB Working Group, "the effectiveness of an oversight system relies upon the consistency of coordination of activities across agencies sharing the oversight responsibility", these discrepancies are a sobering reminder of the potential inadequacies of the current regulatory framework.

Current EPA regulations under the Toxic Substances Control Act (TCSA) likely have a limited impact on the synthetic biology community as most conceivable interpretations of "microorganisms" regulated under the TSCA would exclude "naked DNA", whether in an expressible form or not. Consequently, the regulations would likely only come into force when a functional microorganism (in this case, an infectious agent) is generated. Needless to say, anyone with hostile intentions would be unlikely to pause at this point to file the necessary paperwork. However, as previously noted, the procedures undertaken by the EPA for the protection of confidential business information contained within MCAN or TERA submissions may be of particular relevance to any hypothetical synthetic DNA order screening system, particularly if order histories are to be stored, as per the NSABB recommendations.[240] Conversely, these data storage practices are likely only relevant in an abstract sense as paper (or even CD-based) submissions of nucleotide sequence screening data to a centralized government analysis and storage center would likely prove immensely slow, cumbersome, and expensive, while current EPA practices have little bearing on industry concerns about the security of electronic data transmission and storage.

Numerous commentators have suggested that the US Government implement measures encouraging or requiring Federal grantees and contractors to order only from commercial synthesizers who screen oligonucleotide or gene synthesis orders for sequence homology with known pathogens.[241] These procedures could conceivably be implemented in a manner analogous to the NIH Guidelines. Indeed, some authors have suggested simply extending the NIH Guidelines to embrace these requirements. Other analogues might include current requirements imposed on recipients of NIH or other Federal research grants in regard to human embryonic stem cell research and laboratory animal treatment standards. In the first instance, as has been widely reported, President Bush limited federal funding for stem cell research to projects utilizing cell lines in existence as of August 9, 2001.[242] These restrictions however, are significantly less stringent than those promulgated in the NIH Guidelines in that while the former apply solely to specific research projects (rather than researchers or research institutions) receiving federal funding for human embryonic stem cell research, the latter are imposed upon *all* researchers affiliated with or receiving funding from an institution that has received *any* NIH funding for research involving recombinant DNA.[243, 244] Thus, current restrictions on stem cell research may prove too lax a model upon which to base any proposed regulatory system for synthetic biology. Moreover, to make a model of a policy that is deeply unpopular within the scientific community may prove unwise. Proposals to use

---

[239] ECCN 1c353, "Technical Note"

[240] "Addressing Biosecurity Concerns Related to the Synthesis of Select Agents", National Science Advisory Board for Biosecurity Working Group on Synthetic Genomics, December 2006., p. 13

[241] "Addressing Biosecurity Concerns Related to the Synthesis of Select Agents", National Science Advisory Board for Biosecurity Working Group on Synthetic Genomics, December 2006., p. 11

[242] "Notice of Extended Receipt Date and Supplemental Information Guidance for Applications Requesting Funding that Proposes Research with Human Embryonic Stem Cells

[243] "Federal Funding Regulations on Stem Cells Partition Labs", Wes Allison, *St. Petersburg Times*. July 8, 2007. Available at: http://www.timesargus.com/apps/pbcs.dll/article?AID=/20070708/FEATURES05/707080302/1014/FEATURES05

[244] "NIH Guidelines for Research Involving Recombinant DNA Molecules", April 2002, Section I-C. "General Applicability".

federal funding as an incentive for responsible behavior and a deterrent for irresponsible behavior among the research community have also been criticized as an insufficient response to biosecurity concerns.[245] Thus, it is likely that such proposals will have to be incorporated into a broader biosecurity policy directed at regulating the synthetic biology community.

Another interesting point that distinguishes the regulation of CSNA from the regulation of pathogens (or other dual-use items, as discussed below) is that, with few exceptions, those possessing, producing or transferring CSNA may not know that they are working with dual-use material. This fact has several implications in relation to the regulations discussed. Is a CSNA supplier obliged to determine that the sequences they send overseas are not "associated with pathogenicity" and therefore require review by the Commerce Department? If not, how are these regulations expected ever to be enforced?

Regulations Covering Materials Comparable to Synthetic Oligos

Given that current US laws and regulations such as the Select Agent Rules may have only limited relevance to the production, use, and transfer of CSNA sequences (including those of dangerous pathogens or toxins), the expanding availability of these materials may, as noted by the authors of the Sloan Foundation-funded report on Synthetic Genomics,[246] necessitate the development of new regulations, or at least, the broadening of existing regulations, to address concerns related to their potential misuse. Policymakers may find it instructive to refer to existing regulations on a number of other dual-use materials (such as radioactive substances) as well as to contemporary regulations in force, or under consideration, in other parts of the world. While the following discussion of potentially analogous regulations is far from comprehensive, we hope that it will provide a useful overview of existing regulatory practices, some of which may prove germane to any proposed regulation of CSNA.

Radioactive Materials

The laboratory use of radioactive materials is in many ways analogous to that of CSNA. Both radionuclides and CSNA are widely available laboratory tools and are used by a broadly comparable number of researchers within the United States. Given that many molecular and cell biologists use or have used radionuclides in their work, and that radionuclides are commonly used in more fields than CSNA (such as physics, engineering and physiology) and that radionuclides have uses in many fields outside of the laboratory (such as in agriculture and construction), we can safely assume that the quantity of users of radionuclides is at least as large as the user community of CSNA. However, radionuclides differ from CSNA in a number of crucial ways. Unlike radionuclides, CSNA cannot be considered intrinsically physically dangerous. Instead, they are in some ways analogous to computer code in that their harmful or innocuous nature is determined by their content, and only comes into being when inserted into a system that can act on the information encoded. Consequently, while commercial radionuclide suppliers are "automatically" aware of the potentially hazardous nature of their products, commercial nucleotide synthesizers are most often unaware of the content of theirs. Courier services and customs agents are moreover, theoretically able to detect the presence of radioactivity in an innocuously-labeled package. If one imagines two hypothetical packages, one containing DNA from a hazardous pathogen, the other containing radioactive materials, both labeled as generic "laboratory supplies", sent by well-intentioned, if irresponsible, researchers to their respective international collaborators, it becomes clear that while the second package might very well be recognized as radioactive by a reasonably-alert official at some point in its journey, it is inconceivable that anyone without a priori knowledge of the contents of the first package might notice its potentially hazardous nature. Thus, while safety and security procedures

[245] Garfinkel, M., Endy, D., Eptein, G., Friendman, R. "Synthetic Genomics Options for Governance." Alfred P. Sloan Foundation.
[246] NSABB Report

developed for radioactive materials may prove instructive, there are obvious limits to the applicability of regulations developed for these materials to synthetic nucleotides.

The Energy Reorganization Act of 1974[247] established the US Nuclear Regulatory Commission (NRC), charging it with regulatory oversight over the medical, industrial, and academic use of nuclear materials. The NRC conducts safety evaluations of sealed sources and devices to ascertain their ability to safely contain radioactivity as well as issuing licenses authorizing the manufacture and commercial distribution of products containing radioactive materials. The regulatory responsibilities of the NRC do not include regulation of defense nuclear facilities, which are under the control of the Department of Defense. Perhaps most relevantly, the NRC also controls (via a licensing program) the purchase of radioactive materials for medical, veterinary, industrial, academic/research, and general[248] uses.

Agreement State Program

Under the Atomic Energy Act, states can assume the role of the NRC with respect to the regulation and licensing of radioisotopes, source material (uranium and thorium) and certain special nuclear materials. Authority is transferred via a signed agreement between the Governor of the State and the Chairman of the Commission (as specified in section 274b of the Act). Agreement states are required to maintain NRC standards in all activities. The NRC provides training courses and workshops for agreement states as well as overseeing all state rule changes and regulatory efforts.[249]

While states opting to become Agreement States are not financially supported by the NRC, the collection of application fees helps them to offset the costs of running the program. Becoming an Agreement State allows the states to enjoy greater control over the research, medical, and industrial application of radioactive materials. Thus, in the event that contamination concerns arise, the states are themselves able and (required) to respond. Nonetheless, the NRC retains regulatory control over all power plants, fuel cell facilities, and research reactors. The NRC also regulates all overseas suppliers of radioactive materials. Overall, Agreement States issue 75 percent of licenses, while the NRC administers the remainder.

Evaluation of the Maryland State Program

The application for a new license for the purchase and possession of radioactive materials is fairly simple and straightforward in Maryland. The application itself is a single page in length. The applicant is also asked to submit a single additional page describing in detail the materials desired, the purpose for which they are sought, and the maximum amount of material that will be held in the applicant's facility at any one time. A license is valid for seven years, after which licensees have the option of applying for a license renewal, modifying their applications, or ceasing their usage and possession of radioactive materials. All activities at the end of a seven year license require a visit for verification purposes from the Maryland Department of the Environment (MDE).

While entire research groups, companies, or medical facilities can obtain a license, each license requires a coordinator and a training and safety coordinator for all employees. In an industrial setting, training programs and certificates are required and must be verified before a license is issued. All employees/group members working in facilities with radioactive materials are required to be trained in safety and waste management. In an academic setting, personal references attesting to prior training (in

[247] Available at: http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0980/ml022200075-vol1.pdf#pagemode=bookmarks&page=213

[248] Defined as the "commercial use of nuclear materials by general licensees for activities such as measuring, gauging, or controlling devices, etc." Available at: http://www.nrc.gov/materials/miau/general-use.html

[249] http://www.nrc.gov/about-nrc/state-tribal/agreement-states.html April 30, 2007.

the use of radioactive materials) are required as training in this setting usually occurred in other research groups.[250] After an application has been filed, but prior to the issuance of a license, the facility must undergo an inspection from the MDE. Inspections and audits are generally unannounced and check for cross-contamination issues, waste management, proper labeling and adequate training for all those employees working with radioactive material. The issuance of licenses for certain materials and resources (*e.g.* radiation sources such as food irradiators) require a more extensive background check.

The MDE generally processes applications within seven months of the initial application submission though, typically, licenses are obtained within two months. Maryland currently employs three reviewers under this program. These reviewers report that they enjoy good, open, and well-established communication with the Nuclear Regulatory Commission.[251]

GAO Sting Operation

As reported in the *Washington Post*, a recent (July 2007) GAO sting operation uncovered serious security gaps in NRC licensing procedures. Posing as West Virginia-based (a non-agreement state regulated directly by the NRC) businessmen, GAO investigators managed to obtain portable moisture density gauges containing enough radioactive material (americium-241 and cesium-137) to build a "dirty bomb" qualifying as a "level-3 threat on the International Atomic Energy Agency's scale of 1 to 5, with 1 being the most hazardous."[252] NRC officials reportedly approved the license request submitted by the phony company after a minimal background check that did not include a "face-to-face interview or visit to the purported company" which apparently existed only as a "post-office box at Mail Boxes Etc., a telephone, and a fax machine." In contrast, the GAO investigators were unable to obtain a license in Maryland, which required a site visit by state inspectors and a vetting process that they were told would likely take seven months. In response to the results of the operation, the NRC temporarily suspended the issuance of new licenses pertaining to radiation risks of three or lower and subsequently modified its licensing procedures to require site visits or a face-to-face meeting. Site visits had previously only been required for the issuance of licenses to handle radioactive materials with risk levels of one and two.

Proposed EU Regulations

The European Union Justice Commission, issued a "Green Paper on Bio-Preparedness" seeking proposals from community stakeholders to address the risks of misuse of biotechnology on July 11, 2007 (the text of the proposals was drafted prior to the discovery of the terrorists practicing as doctors in the UK).[253] Although the drafters of the Green Paper do not explicitly consider risks of misuse of CSNA, some of the proposals mooted are of interest for our work. One proposal in the Green Paper describes an EU-wide scientific licensing regime: "European rules for national certification and registering of facilities with regard to compliance with bio-standards and the credentials and competences required of researchers" and "A European…system for certifying reliable and trusted facilities and researchers, facilitating the secure and safe exchange of samples and sensitive research results."[254] Another proposal describes a regime similar to the NIH-guidelines: "Research grants should not only be conditioned upon the quality of a proposal, but also upon the ability of the given applicant to comply with the bio-standards as well as

[250] Interview with Nathaniel Wrutsky, Maryland Department of the Environment – Radiological Health, June 29, 2007, although personal experience of this research group suggests that, in practice, academic institutions require re-training when new investigators arrive.
[251] Interview with Nathaniel Wrutsky, Maryland Department of the Environment – Radiological Health, June 29, 2007.
[252] Day, K. "Sting Reveals Security Gap at Nuclear Agency", *Washington Post*, July 12, 2007.
[253] Commission of the European Communities, "Green Paper on Bio-Preparedness" COM(2007)399 final, July 11, 2007. Available at: http://www.europa.eu/rapid/pressReleasesAction.do?reference=MEMO/07/289&format=PDF&aged=0&language=EN&guiLanguage=en
[254] Commission of the European Communities, July 11, 2007.

possible future security guidelines." Comments from academe and industry will be sought until October of 2007, after which the European Commission will likely draft a White Paper laying out recommendations for EU-wide biosecurity legislation. At this stage, it is difficult to ascertain the probable contents of these final recommendations.

Conclusions

Although the SAR provide a robust framework for the control of the pathogens themselves, at present, they are generally understood not to apply (except for variola virus) to fragments of a genome of a pathogen or fragments of a gene encoding a toxin, or potentially even to an intact genome of viruses that need more components to enable expression. Furthermore, the quantity of researchers (approximately 8,000) regulated by the SAR is far fewer than those ordering CSNA (~200,000), complicating the direct application of procedures from the SAR to this field. EPA regulations under TSCA do not apply to naked nucleic acid, but would apply when the nucleic acid is inserted into an organism. Where regulations under TSCA are relevant are the mechanisms used to control and protect confidential business information on biologicals. Commerce Control Regulations *should* apply to a significant fraction of the several million CSNA orders exported yearly, but, in practice, CSNA companies do not apply for export licenses for several reasons: BIS lacks the capacity to vet these orders, CSNA companies are largely ignorant of these regulations, the time required to obtain a license is far greater than the usual delivery time of CSNA products, CSNA suppliers may not know their product is dual-use and, due to the difficulty of properly identifying CSNA products in a package, it is impossible to enforce.

Because the quantities of researchers using radionuclides and CSNA are roughly similar, mechanisms to license and regulate the use of radionuclides may be instructive to the oversight of CSNA. Robust oversight of radionuclides clearly requires on-site visits with potential licensees. The EU is now considering licensing life-scientists to prevent hostile actors gaining access to useful reagents and know how.

Chapter 5: Options for a CSNA Order Screening System

Components of a Notional System Overview

In order to facilitate a discussion about CSNA screening options and to highlight decisions that must be made, we present a notional system for oversight of the CSNA industry. The notional system, presented as a flowchart below (Figure 22), includes steps that may not be necessary in all screening systems. As an order is received, it will pass through the system in the direction of the arrows. Below we will briefly walk through each step in the system highlighting when decisions must be made when choosing a CSNA screening system. In the following sections we will present an in-depth look at the decisions that must be made for each step and the options that should be considered. In some cases, to avoid unnecessary repetition, the sections below are organized by the parameters that underlie the trade-offs among options, in other cases, by the options themselves. Steps that are self-evident are not discussed so not all numbered steps in the system will be represented in the text.

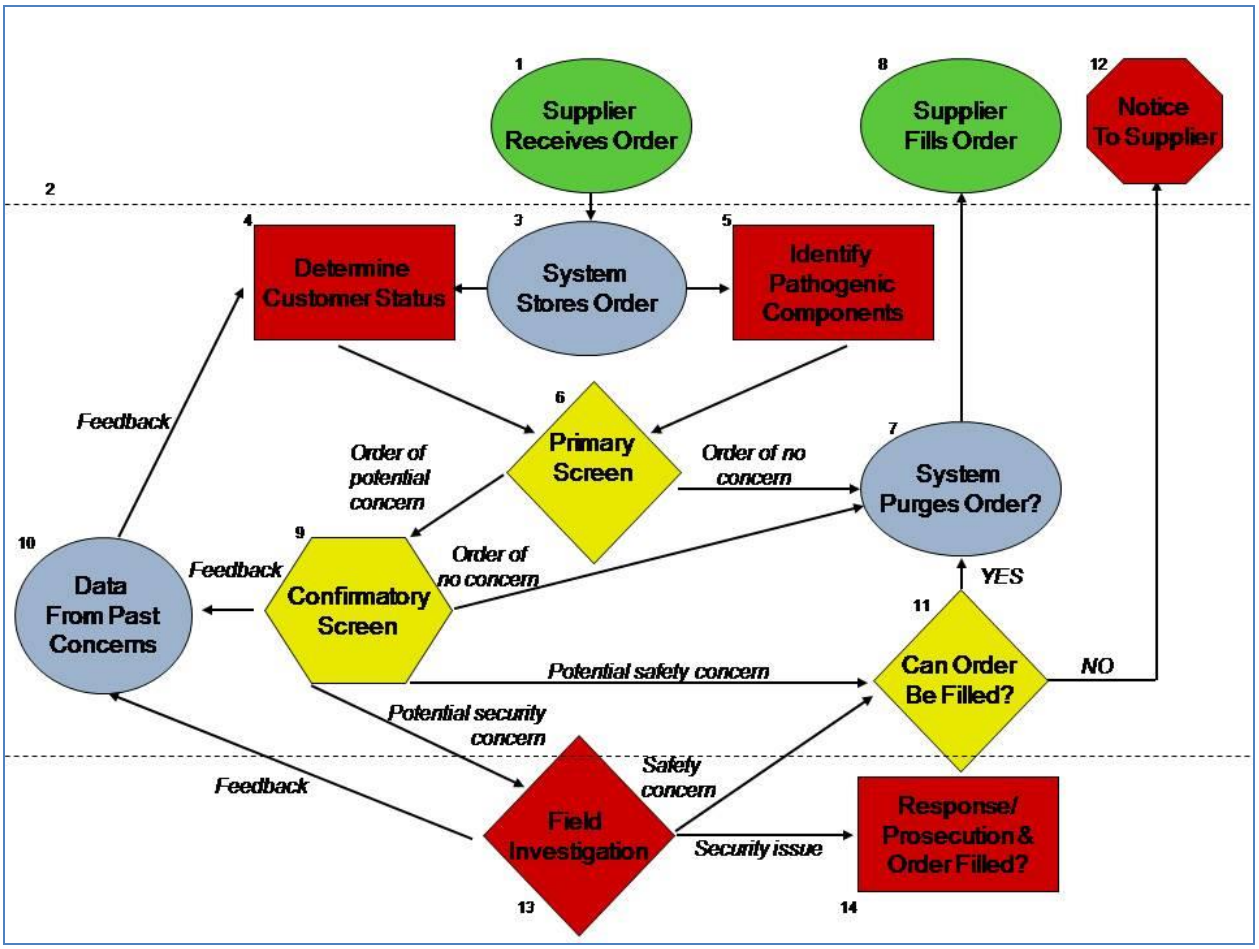


Figure 22. Flow chart representing a notional system for the oversight of the CSNA industry. Not all steps depicted may enter into the final system.

1. **Supplier Receives Order:** Our system begins with CSNA suppliers receiving orders, which can be received via the company's webpage, e-mail, telephone or fax. One of the first decisions that must be made when choosing a CSNA screening system is which products (genes, oligos, etc.) should enter the system and which, if any, should be exempt from oversight.

2. **Lines of Control:** The notional system includes two dotted lines. Decisions must be made about who should perform the steps appearing between the two lines. Steps appearing above the upper dotted line will be performed by the CSNA providers while those appearing below the lower dotted line will be performed by a government-run organization.

3. **System Stores Order:** CSNA orders that will be screened must be entered electronically into a screening system which will be used in Step 4 (determine customer status) and if applicable Step 5 (identify pathogenic components).

4. **Determine Customer Status:** Customers wishing to order CSNA will be evaluated to identify the security and/or safety risks associated with their purchase. This step is required whether or not pathogenic components are screened because the quantity of legitimate orders for CSNA based on pathogens will be significant.

5. **Identify Pathogenic Components:** If a screening program is chosen that incorporates sequence screening, this step will identify sequences of concern. If sequence screening is not included in a screening system this step, will not be performed.

6. **Primary Screen:** The results of Step 4 and, if applicable, Step 5 are reviewed to identify the security and/or safety risks involved in the purchase. Orders of concern will proceed to Step 9 (confirmatory screen) while those of no concern will proceed to Step 7 (system purges order) and then be filled in Step 8.

7. **System Purges Order?**: A decision must be made as to whether to store or purge the information used to determine the safety of an order.

8. **Supplier Fills Order:** CSNA orders that are identified as safe are filled and shipped to customers.

9. **Confirmatory Screen:** If the results of the primary screen (Step 6) indicate an order is of potential concern (for instance the customer's legitimacy can not be confirmed or a sequence is identified as potentially dangerous) further inquiry into the order will be performed to determine whether or not the order constitutes a safety or security concern. Orders determined to be safe will proceed to Step 7 (system purges order?) and be filled in Step 8. While those that may present a potential safety concern will proceed to Step 11 (can order be filled?) and orders that present a security concern will proceed to Step 12 (field investigation), respectively.

10. **Data from Past Concerns:** Results from confirmatory screens (Step 9) and field investigations (Step 13) are collected to help improve the screening system. This feedback may prevent false hits from future primary customer screens (Step 4).

11. **Can Orders Be Filled?:** If the confirmatory screen (Step 9) indicates there is a potential safety concern, a decision must be made as to whether the safety concern can be addressed and the order filled (proceed to Step 7), or if the safety concern is serious enough to deny filling the order (Step 12).

12. **Notice to Suppliers:** If the field investigation determines that the order should not be filled, the supplier is told not to fill the order. We note that in some cases a supplier may be told to fill an order even if a customer poses a potential security risk to aid an ongoing investigation.

13. **Field Investigation:** If the confirmatory screen (Step 9) identifies a potential security risk, a field investigation is launched by a government agency to determine if there is in fact a security issue**.** Those orders determined not to pose a security risk by the investigation will proceed to Step 11 where a determination can be made whether or not to fill the order.

14. **Response/Prosecution & Order Filled?:** Field investigations that uncover security concerns may lead to further action by the government including denying orders from being filled and/or prosecution of those placing the order.

Phased Implementation and Tailoring the System

As the following sector is considered, one should not infer that in order to reduce risks of misuse of the CSNA industry, policymakers must arrive at one system that is designed to either include or completely exclude certain components of the industry. It is possible that the CSNA oversight system could include one set of mechanisms for the oversight of the manufacture of genes or long oligos and another system for the oversight of short oligos. For instance, due to the volume of orders and issues related to false positives, it is possible that the right system for the regulation of the CSNA system as a whole may involve customer screening and sequence screening for gene and long oligo orders, but simple licensing for short oligos.

Furthermore, because of the differences that exist in various segments of the CSNA industry, it would be possible to phase in the implementation of oversight of the entire CSNA industry by implementing oversight of one segment at a time. For example, because of the gene industry's relatively low volume of orders and quantity of customers, its relaxed time constraints, the fact that the bulk of the companies in the industry are taking a proactive biosecurity/biosafety posture, and the possibility that their products may be the most enabling, the synthetic gene industry could be the first to adopt an oversight regime. Once the oversight is proven to be effective, not unduly burdensome, and initial problems with implementation are solved, oversight could be extended to the oligo industry (and perhaps started with oligos greater than 35nt long).

What Will Enter the System

When choosing a regulatory system for CSNA, one of the first questions that should be considered is: what types of CSNA should enter into the screening system in the first place? One option to consider is the regulation of all commercially sold CSNA. This would include synthetic genes of all sizes, DNA oligos of all sizes, DNA-analogs such as synthetic RNA, morpholinos, peptide-nucleic acids or locked-nucleic acids, modified oligos such as those that have been biotinylated, or phosphorylated as well as oligos that are in non-standard formats (those that are attached to beads or a microarray matrix, as well as those that can only be purchased in large scale formats (mg quantities or greater). Other options for regulation might exclude some of the CSNA groups described above. One must consider whether or not to regulate oligos or a subpopulation of oligos such as those that are less than 35nt in length, DNA analogues, modified oligos, or those that are sold in non-standard formats. Below, we will discuss the costs and benefits that must be considered when choosing regulation in relation to its effect on biosafety, the ability to catch hostile actors attempting to construct a synthetic virus, the ability to catch hostile actors attempting to PCR pathogenic genes from environmental samples, as well as the quantity of

synthetic nucleotides that would be screened (and the cost associated with screening) and the quantity of companies that would be involved.

Preventing Viral Synthesis by Hostile Actors

One major concern surrounding the synthetic nucleotide industry is that CSNA (in the form of genes or DNA oligos) could be used by hostile actors to construct a virus (or directly build a gene for a pathogenic component for expression in bacteria) for use in a biological attack. Although screening against all CSNA sequences provides no guarantee that CSNA could not be used for such activities, it would be the most comprehensive screening system and therefore may be the best defense against nefarious viral synthesis. Another possibility to consider is the screening of only those oligos greater than 35nt in length and genes. As previously discussed in Chapter 3, it would be unlikely that oligos less than 35nt in length would be used to construct a synthetic virus (or a gene) de novo. Because the majority of oligos ordered today are shorter than 35nt, excluding these short oligos from the system would prevent about 80 percent of the orders from entering the system, reducing operating costs, false hits and burden on the industry and researchers. False hits would be further reduced because short oligos have a greater probability of serendipitously matching a pathogenic sequence than longer oligos.

Likewise, one could consider excluding DNA analogs, modified nucleotides and other non-standard format nucleotides from screening. Many DNA oligo modifications, DNA analogs, and other non-standard formats (such as those attached to a chip) are unsuited for direct use in gene (and therefore, genome) fabrication. However, many of these modified oligos, RNA oligos and some DNA-analog hybrid molecules could be modified via enzymatic action or restriction endonucleases to result in unmodified DNA oligos perfectly suitable for gene fabrication. It should be noted that the only reason to use DNA analogs or modified bases in gene fabrication is to avoid detection by a screening system (scientists use DNA oligos to build a copy of the genome of even RNA viruses). Each one of these non-standard CSNA types has a disadvantage in comparison to unmodified DNA oligos when considering the complexity and likelihood of success of the protocol to arrive at the desired products. Therefore, the inclusion of these molecules in a screening system only makes sense if one believes that the adversary is aware of (and feels confident in the abilities of) screening systems. Similarly, oligos that are only sold in large (mg or greater) quantities are 100 fold more expensive than standard quantity oligos, and these quantities are completely unnecessary to produce genes or viral genomes through annealing and ligation. Conversely, the quantity of oligos needed to create a virus or directly construct a gene, combined with the cost of oligos bought in milligram quantities, may be prohibitive to some actors. For those adversaries that could afford oligos bought in excessive quantities, the costs associated with using oligos sold in large quantities only makes sense if they are trying to avoid detection by the system.

Lastly, one could consider screening only genes. Excluding oligos would provide much less protection against viral synthesis, as hostile actors may be able to construct a virus using unregulated synthetic DNA using the same techniques Eckard Wimmer used to construct the Polio virus. This approach, while providing limited utility in the prevention of hostile actors synthesizing viruses, would be the least costly due to the relatively low volume of gene orders compared to oligo orders. Furthermore, the length of genes (and the detailed information collected about genes by some companies before synthesis) will greatly reduce the false hit rates.

Preventing PCR Amplification of Pathogenic Genes by Hostile Actors

Although much of the focus of regulation of CSNA has been placed on the prevention of the synthesis of viruses de novo, CSNA may also serve as primers for the amplification of toxins or virulence genes from environmental samples. Most PCR primers are less than 30nt in length; therefore a major disadvantage associated with excluding CSNA products less than 30nt is the inability to screen PCR primers. Screening

all CSNA will, of course, be the most effective way of preventing a CSNA order that could be used as primers for PCR amplification of pathogenic genes. As discussed above, one could consider excluding DNA analogs, modified nucleotides and other non-standard format nucleotides from screening as these are unsuited for direct use in a PCR reaction. However, with few exceptions, we cannot say with certainty that any type of modification renders an oligo completely unable to be used in a PCR reaction (especially after the possible removal of these modifications by enzymes is considered). Furthermore, exclusion of oligos that are only sold in large (mg or greater) quantities is of limited benefit, as PCR requires only two oligos and therefore the increased cost of ordering these oligos in large quantity would be modest.

Biosafety

The previous two subsections have focused on regulating CSNA to prevent nefarious activity; however, the issue of biosafety should not be overlooked. One potential concern that should be considered is the improper and unsafe use of CSNA by legitimate researchers. A regulatory system that screened all CSNA could help identify harmful sequences and ensure customers understood the associated risks. As discussed above, a screening system that screened only genes would be less effective as oligos could be used to construct viruses or to amplify genes from environmental samples. Because DNA analogs, modified CSNA (except for degenerate primers and oligos with universal bases) and nonstandard CSNA formats are not ideally suited for these applications, they would probably not be used by legitimate researchers who are not trying to hide their activities. Therefore inclusion of non-standard CSNA orders would not likely need to be taken into account when considering regulations that would promote biosaftey.

Order Volume/Screening/Cost

From the information presented above, it is clear that screening all CSNA orders will provide the best protection against hostile actors obtaining CSNA that could be used for nefarious purposes as well as the best protection against accidents associated with CSNA. Despite this fact, the costs associated with screening all CSNA must be weighed against the benefits that can be obtained by screening only a subpopulation of synthetic nucleotide orders. As presented in Chapter 3, approximately 30 million oligos are synthesized worldwide each year (most of which are produced in the United States). Of these orders greater than 80 percent are less than 35nt in length. By screening only CSNA 35nt or longer (those sequences that would be long enough to be used in synthetic genome manufacturing) the total number of orders that would require screening would decrease by 80 percent thus reducing the resources associated with screening and the resources needed to investigate false hits. Furthermore, including only longer oligos and genes in a screening system would further reduce the false hit rate because smaller oligos are more likely to match “restricted” sequences by chance. Additionally, because the majority of oligo users order only oligos less than 35nt in length, fewer customers would be subject to screening. Resources devoted to screening could be reduced further if only genes were subject to screening regulations as only about 50,000 genes are currently synthesized each year.

Conversely, few benefits would be realized by excluding modified nucleotides or non-standard format nucleotides from oversight as these products make up only a small fraction of the synthetic nucleotide industry. However, the benefit of excluding these orders is marginal when one only considers oligo volume, but it would leave some smaller CSNA companies (for whom screening may be burdensome) outside of the system.

Quantity of Companies Involved

As mentioned in Chapter 3, Gryphon Scientific has identified 29 US companies currently producing oligos and 20 US companies that currently make synthetic genes (note that nine companies are counted in both groups as they produce both products.) Regulation of all CSNA would require that all 40 companies

participate in regulatory actions. By choosing to regulate only the synthetic gene industry, the quantity of companies involved that would need oversight would be cut in half. With regards to the number of CSNA producers involved, there does not appear to be an advantage associated with screening only oligos 35nt or greater as all oligo companies we spoke with made at least some products greater than 35nt. However, the number of companies involved could be decreased by excluding DNA analogs, modified nucleotides and non-standard-format oligos because the oligo business has become extremely specialized and many companies do not make standard-format, unmodified DNA oligos. It should also be said that because the companies that make only non-standard CSNA are generally smaller than those that make unmodified, DNA oligos, they are less able to bear any additional regulatory burden.

Table 4. Stoplight chart comparing the effectiveness of including various types of CSNA in an oversight system with regard to cost, safety and security issues.

Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given category

Parameter	Everything	Genes Only	Everything Except			
			Oligos <35nt	DNA Analogs	Modified Nucleotides	Nonstandard Format Nucleotides
Detect illicit de novo viral synthesis	🟢	🔴	🟢	🟡	🟡	🟡
Detect illicit PCR of pathogenic genes from environmental sample	🟢	🔴	🔴	🟡	🟡	🟡
Benefit to biosafety	🟢	🔴	🟡	🟢	🟢	🟢
Order volume/screening/cost	🔴	🟢	🟡	🔴	🔴	🔴
Quantity of companies involved	🔴	🟢	🔴	🟡	🟡	🟡

Who Will Provide Oversight?

Once a determination is made regarding which components of the CSNA industry are subject to oversight, one must decide which entity is best (government, industry or a third party) to provide that oversight. The following section will discuss a few considerations that should be taken into account when determining which entity is best to oversee the industry.

Compliance

As with any regulatory requirement, one must consider how to evaluate performance and ensure compliance. Certainly the most effective way to obtain regulatory compliance in the CSNA industry is to require that screening of orders be conducted by the federal government. Although industry-regulated screening would rely more heavily on the honor system, substantial compliance may still be attained. Audits and inspections by government agencies could be used to assess compliance and correct overly lax screening done by industry groups. This type of auditing may include scrutiny of a record of orders that failed a primary screen and then passed a secondary screen, for example. No matter how it is achieved, monitoring of compliance will require at least some cost to be borne by the government even if all screening is performed by industry. Metrics to assess risk reduction will require performance of a full weapons-production assessment.

Consistency/Gaming the System

Currently there is no consistency in sequence screening among those who provide CSNA, largely because their efforts are not coordinated between companies. While some companies consistently screen customers and sequences, others have more lax screening processes including those companies that do no screening at all. While it would be possible to enact regulations requiring all US companies to screen sequences, without specifying what needs to be screened and how, it would be difficult to maintain a consistent screening practice in all companies. A lack of consistency could enable hostile actors to "game the system" by finding the company with the weakest screening system to supply them with needed CSNA. To eliminate this problem, companies could be required to use a standard screening program (dictated by the government) and instructions on how to proceed in the event a customer or sequence is recognized as a potential risk. This approach would allow the screening of sequences locally while maintaining consistent screening across all companies, thus eliminating the problem of a "weakest link" that could be exploited by an adversary. Although consistent local screening could eliminate one potential for gaming the system, it would still allow other means of evading detection, such as splitting an order between several companies to avoid detection. This problem could only be avoided by establishing a centralized (government-run) screening system where sequences would be screened (and potentially archived) together regardless of which company the order came from. Although a single set of standards would make it easier for an adversary to determine which actions are taken to screen orders system-wide, an adversary really only needs to understand the practices by one company in order to determine the best means of ordering CSNA without detection. Therefore, centralized screening has no significant disadvantage compared to a system wherein each company sets their own screening standards simply because it uses a single screening system.

Competitive Advantage

A system that is run by the federal government or standardized across companies has yet another advantage: no one company could gain a competitive advantage over others by performing a cheaper, yet less effective, screen. If companies are allowed to self-regulate (using their own screening software), it becomes advantageous for a company to use a less stringent screening system, because fewer orders will require the resources of a secondary screen, and fewer orders will be rejected. Conversely, companies that choose rigorous screening systems may engender delays processing and shipping orders and/or lost orders due to the identification of a legitimate order as one that is suspect. Either possibility may lead customers from companies with more stringent screening systems to companies with more lax ones.

Time Delay

Although sequence monitoring by the federal government, rather than industry, will help ensure compliance and consistency, the potential downfalls of such a system cannot be ignored. When considering possible regulatory action one must consider how regulations will affect the ability of researchers to obtain CSNA in a timely manner. Although orders for synthetic genes typically take at least one week to synthesize and ship, many oligo companies offer same day synthesis and next day delivery of custom made oligos. At this point, it is important to recall that many customers of the CSNA industry we interviewed thought that rapid receipt of their oligos was essential, and that delays of a day or so would cause them to look for other suppliers. The CSNA suppliers understand this sentiment. Throughout the course of our industry interviews, oligo providers repeatedly expressed that a delay in synthesizing and/or shipping products to their customers would destroy the US oligo industry and would lead to the relocation of companies outside of the country in an attempt to avoid what the industry feels is an unacceptable competitive disadvantage. As previously noted, this assessment by the industry may not be the result of well-thought-out cost/benefit calculations. Given that we estimate that 30 million CSNA orders are placed

each year, screening of each of these sequences by an outside government organization would most likely lead to delays in the processing, synthesizing and shipping of CSNA.

Even if enough resources were devoted to screening at the national level to eliminate a backlog, delays in sending orders and screening information between suppliers and screeners are likely to cause waste (the production of orders that should not be sent) and mistakes (orders are produced and shipped before the notice to stop production is received). Many CSNA providers (including oligo providers) have indicated that they have at least several minutes between when an order is received and when synthesis is begun during which time orders could be screened. By allowing industry to regulate the screening of CSNA orders, the delay in order processing, synthesis and shipping could be kept to a minimum. Because the vast majority of CSNA are ordered for legitimate use by researchers, and these researchers claim that delays in oligo delivery would be a "deal breaker", the costs associated with delaying oligo synthesis and shipment must be considered.

A hybrid oversight system involving government standards and government approved software, but where screening occurs at the supplier may help decrease the time delay while maintaining consistent screening.

Protecting Proprietary Information

Many CSNA industry representatives we spoke with expressed a concern that if screening of orders (in this case, the content of the order itself, not just information on the customer) was to be conducted off-site that the confidentiality surrounding the sequences would be compromised, or at least their customers would perceive a loss of confidentiality. Several companies claimed that allowing sequence information to be sent (to the government) outside of their firewall would lead to contract violations and lost customers, although we note that binding laws and regulations override contract provisions. Others opined that clients would choose to synthesis their own CSNA rather than risk the compromise of the confidentiality of their research. By allowing providers of CSNA to perform sequence screening locally, the issues surrounding confidentiality of proprietary information could be avoided. Likewise, offsite screening of an order that does not include the sequence (i.e. customer screening) may also solve the problem of sequence confidentiality. A hybrid approach, where only orders that failed an initial screen conducted by the suppliers would be sent to the government for review, would allow most orders to remain confidential.

Cost

Costs to the federal government will be kept lower if providers of CSNA are allowed to screen orders internally. Although local screening will not eliminate all costs to the government, as the government will need to monitor industry to ensure compliance and provide support in the event a threat is identified, industry will absorb costs related to compiling a list of sequences to screen against, a list of suspicious customer traits, and developing (or leveraging existing) screening software as well as the operating cost (as measured by employee time expenditure).

A national system would require that the government compile a list of sequences to screen against, the list of suspicious customer traits, and develop (or purchase) screening software as well as employ enough personnel to screen millions of sequences a year. Although it may appear this system would relieve the burden on screening from industry, the predicted customer loss due to confidentiality issues and increase in processing delays would likely lead to costs for industry as well, and probably underlies their resistance to this type of system.

The most cost-effective system for both industry and government might be one in which the government identified which sequences should be screened against, identified the customer traits that are suspicious, and developed or recommended existing screening software programs while the screening is actually performed by each company. Development of such a standardized system would help to distribute the costs associated with screening between the federal government and the providers of CSNA. This hybrid system would likely reduce the costs to industry associated with lost customers due to time and confidentiality issues while significantly decreasing the federal government's personnel costs.

Industry Backlash

Although many *gene* companies we spoke with indicated that they currently screen orders (via customer or sequence screening), very few of the *oligo* companies we spoke with screen orders, and of those that screen, none currently screen sequences (all screen customers only). As mentioned in Chapter 3, our contacts in the oligo industry have expressed concern that a government-run CSNA screening system would lead to the downfall of the US oligo industry. Likewise, both the oligo and gene companies we spoke with expressed concerns related to the loss of proprietary information if order screening was to be conducted by a third-party screener. Although any mandatory screening will likely not be well received by those in the oligo industry, based on our interviews, it is clear that a screening system that will allow for local screening will be received with less opposition than one that would require offsite, third-party screening. While a few in the gene industry use there own screening systems that are integrated into their gene design system and would therefore be opposed to using government supplied software, most oligo companies liked the idea of screening software being provided for them. Many gene synthesis providers also indicated that they would be in favor of having a standard set of screening rules to follow, especially information about what sequences may be considered "dangerous" (even those that would prefer to use their own software to screen). This observation reinforces that a phased approach, beginning with gene sequences may be appropriate.

Improving the System

As the system operates, undoubtedly information will be collected that will enable improvements to be made in its performance. Factors contributing to false hits at the primary or secondary screening can be logged and accounted for in each step. Furthermore, data related to the identification of suspicious orders could be logged and included in the system to ensure that an adversary cannot defeat the system. If the system is run by a centralized authority (either the government or a third party), this information can be processed and included in the operation of the system in real time. For instance, if a suspicious person attempts to obtain a particular sequence associated with a pathogen and is denied, the system could decrease the threshold for screening related to that pathogen for the near future. As another example, if a company denied a suspicious order from a particular customer in an unusual country, the system could decrease the alarm threshold for all orders from that country for the near future.

This type of information sharing would be greatly complicated by a system in which screening is done locally. Real time improvement of the system would only be possible if data on system hits were shared quickly and the compensating changes to the system were disseminated back to the users and adopted quickly. Clearly, no matter who is performing screening, system improvements based on experience with operation of the system are an essential part of any deployment scheme, and the lessons learned from system operators should be collected frequently to ensure the system is as useful and efficient as possible

Conclusions

While there does not appear to be one screening system that fits all needs, when considering the options for regulation of the CSNA industry it is important to remember that there does not need to be one system

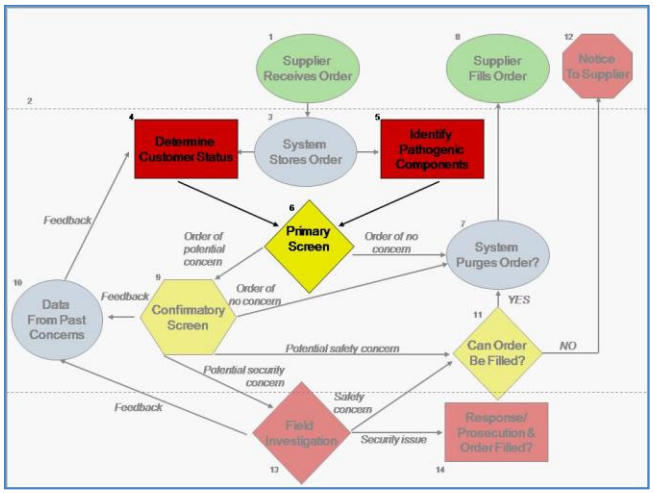
that governs the entire CSNA industry. Given the diverse nature of the CSNA industry, the best system may involve a different set of standards for oligo providers than it uses for gene providers. The trade-offs are displayed in Table 5.

Table 5. Stoplight chart comparing three options for control of a screening system.
Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given category

Parameter	Federal Screening	Local Screening	Federally Standardized Local Screening
Compliance	🟢	🟡	🟡
Consistency/Gaming system			
Figuring out standards	🟡	🟡	🟡
Ordering from multiple companies	🟢	🔴	🔴
Finding weak screening	🟢	🔴	🟢
Competitiveness issues	🟢	🔴	🟢
Time delay	🔴	🟢	🟢
Propriety issues	🔴	🟢	🟢
Cost to industry	🟢	🔴	🟡
Cost to government	🔴	🟢	🟡
Distribution of effort	🔴	🟡	🟢
Improving system	🟢	🔴	🔴
Industry backlash	🔴	🟡	🟡

Steps 4, 5, and 6: Primary Screen: Customers and Sequences

In the previous two sections, we discussed the options of what types of CSNA may enter an oversight system and who would provide the oversight. Next, one must decide how this oversight should be implemented. In this section, we discuss the options for the initial screening of CSNA orders. One option is to screen customers who wish to order regulated CSNA, while another option is to screen both CSNA sequences and customers. Given that there are plenty of reasons to order CSNA that may be regulated (like sequences from select agents), we feel that screening orders in the absence of any screening of the customers is untenable as long as there exists a legitimate use of CSNA related to even the most dangerous pathogens.



Customer Screening

Several options are available for screening customers including verification of the customer's e-mail and shipping address, investigating a customer's credit history, comparing the names of potential customers against those on the denied parties list or against a counter terrorism list and the licensing of all CSNA customers (a pre-screen). When used in combination with sequence screening, more customer screening options become available. Furthermore, several customer screening options could be adopted together (such as e-mail validation combined with screening against known terrorist and known proliferator lists). In this section, we describe the strengths and weaknesses of various options for screening customers.

Before we proceed, some have suggested that criminal background checks, such as those performed by employers or gun merchants may prove an appropriate screening mechanism for CSNA customers. Others have suggested that customer screens should be employed to ensure that those who have been involuntarily committed cannot obtain CSNAs. Ultimately, however, the security concerns raised by firearms and CSNAs are vastly different, such that the "targets" of customer screening are barely comparable. For example, firearms are often misused by criminals, and mentally unstable individuals have no issues correctly operating a firearm. Neither criminals nor the insane are likely to have the training (or perhaps the patience) to misuse CSNA, and the cost of performing this type of customer screen will probably have little tangible benefit.

The American Type Culture Collection takes several steps to screen their customers, asking questions about the ultimate use of the culture, the types of equipment to analyze and contain the organism, the background of investigators who will work on the organism, and other data related to the customer's practices and capabilities. That being said, unless the customer is ordering a select agent, the ATCC takes few steps to verify that the answers given by the customer are accurate or truthful. That is, the identity of the biosafety officer, the presence of laboratory safety equipment and the purpose of the research could be fabricated by the customer in order to obtain organisms of interest. Because the transfer of select agents (presumably, of greater interest to hostile actors than other organisms) is governed by a regime that does have significant verification steps, the primary customer screening role of the ATCC for the transfer of non-select agents is to ensure biosafety, and therefore verification may be considered unnecessary. In this section, we therefore consider which customer screening options require further validation and which data on the customer are inherently reliable in itself.

E-mail/Shipping Address Screening

One simple method that can be used to screen customers wishing to order CSNA is e-mail and shipping address screening. This is the method adopted by oligo producers who perform screening for security purposes (as compared to ensuring that they get paid for their efforts). At a minimum, this method would prohibit shipment of CSNA to residential addresses and to those individuals without an institutional e-mail domain name (for example those placing orders with a hotmail or Gmail addresses). This screening could be performed onsite or by a third-party federal screener and would be relatively inexpensive. More extensive screening of this type involves the verification of customer names and company addresses using company directories and internet searchers. This method would be minimally burdensome to researchers as screening would not likely cause a delay in processing or shipping speed for the vast majority of legitimate orders and would not require any sort of registration of the customers. Although this screening method could be implemented quickly (as this information is already available to CSNA providers) its effectiveness at increasing CSNA biosafety would be minimal, because well-meaning researchers would be granted access to any CSNA, even those that they may not be trained to handle safely. Further, its benefit to biosecurity is dubious; hostile actors wishing to evade detection could setup a sham company. Unless this preliminary screen flags orders from all small biotechnology companies, sham companies would likely pass through this type of screen as well.

Credit History

An investigation into a CSNA customer's credit history could be used as a screening mechanism. This method, which could be implemented locally or offsite by a government-run screening program would require that all customers' affiliated companies be screened using a system, such as that run by Dun and Bradstreet, which allows access to a company's credit history. Individuals not affiliated with a company possessing past credit history would be flagged. Much like e-mail/shipping address screening, verification of a customer credit history would be minimally burdensome to researchers since this type of screening is already performed and would not likely cause a delay in processing/shipping speed. Furthermore, the operational burden of this system would decrease after implementation as more and more customers would be registered in an "approved" database. Although this screening system would be slightly more expensive than screening using shipping and e-mail addresses alone, because most companies that provide credit verification programs charge a fee (~ $2,000/ year), the cost to operate this screening system would still be relatively low. Furthermore, many of the CSNA companies we spoke with already use a credit verification program to help prevent synthesis and shipment of CSNA to customers who are unable to pay. Furthermore, introducing this practice to companies that do not currently perform credit verification could be accomplished relatively quickly thereby allowing prompt implementation of this screening regulation. Although this effort is slightly more effective than shipping/e-mailing verification in providing a modicum of biosecurity benefits (hostile actors wishing to gain access to CSNA would need to setup a sham company with a credit history), this method provides no biosafety benefits.

Although it is beyond the scope of this report to consider internationalizing the system, at this stage it is worth considering that many customers of US-based CSNA producers are foreign. For foreign companies that do frequent business with the US, obtaining a reliable credit check is trivial. However, it may be difficult to obtain the credit histories of small and new companies established overseas, complicating the use of this mechanism to screen these customers. It should be noted that small businesses and foreign institutions both represent only a minority of the customers of the CSNA industry, and their loss may be considered worth suffering in the face of more burdensome oversight regimes. One cost of such action would be the creation a small but potentially lucrative market outside of the system for legitimate customers that could be exploited by those with malevolent intent.

Denied Parties List

The Denied Parties List is a publicly available database maintained by the Bureau of Industry and Security of the Department of Commerce. It contains the names of individuals and organizations linked to weapons proliferation activities, both within the United States and abroad. Exporters or re-exporters of dual-use materials are required to check against the list for all parties to their proposed transaction (including freight forwarders, intermediate consignees, and the ultimate consignees) against the most recent version of the list. While the Department of Commerce only regulated exports and therefore does not require that companies screen their domestic customers against the list, it does recommend that they do so, to avoid unwittingly passing on sensitive technology or materials to US residents known to be involved in proliferation activities.

There are also several commercially available proliferation-prevention screening databases, such as the "Risk Report" produced by the University of Wisconsin Project on Nuclear Arms Control. The Risk Report is developed from unclassified information gathered from government documents (including the Denied Parties List), manufacturers' brochures and websites, tradeshow directories, industry databases, and "credible" media reports on persons and entities around the world suspected of ties to proliferation activities. According to the Wisconsin Project, many governments across the world including Canada,

France, Germany, Japan, the United Kingdom, and the United States as well as leading multi-national corporations subscribe to the Risk Report, utilizing the database to investigate suspicious outbound shipments.[255] The database is available via the web, where it is updated every three weeks, or on CD-ROM, updated bimonthly. Single-user subscriptions to the Risk Report are currently available for $2,800 per year.[256]

Much like e-mail/shipping address screening and credit-check screening, screening against the Denied Parties List or with the Wisconsin Project's Risk Report software would be likely to have a minimal effect upon the legitimate scientific research community, as customers would only be screened against known or suspected proliferators. Perhaps the only concern such screening would raise with researchers would be the possibility of false matches based upon shared or similar names. However, this is unlikely to prove to be a major hurdle and could be easily overcome by a simple verification step such as a phone call to a customer's department or institutional biosafety officer. Proliferation-prevention screening could also be implemented almost immediately, at a relatively low cost. However, while such screening offers some biosecurity benefits, it likely will not capture orders from terrorists not known to be connected to proliferation activities. Moreover, proliferation-prevention screening is unlikely to offer any biosafety benefits whatsoever.

Counter Terrorism List

The TSA currently maintains a No-Fly List, which contains the names of individuals who are not permitted to board a commercial aircraft for travel in the United States. A similar counter terrorism list could be employed to prevent the sale of CSNA to known or suspected hostile actors. As a notional concept of operations, the screener would maintain a database of customers who have been previously screened and had "passed" the screen. When an order is received from a customer who has previously been screened, no further screening is necessary. If the customer is new to the system, the screener would enter the name and institution of the customer into a website, to which the website would return a rating of pass or fail. If this is the concept of operations that is implemented, orders that are not flagged by the system would be cleared quite rapidly.

Unlike the customer screening methods described above, a counter-terrorism list would likely require some federal involvement in the screening processes as such a list would likely not be open source. Because this option requires a mechanism for the constant communication between CSNA providers and the federal government, this screening method would be comparatively slow to implement. Likewise, unless the counter-terrorism list is identical to other counter-terrorism lists established for other purposes (like the No-Fly List), establishing and maintaining a list and federal personnel to perform the screening would be an additional expense.

Although this mechanism for screening would not likely delay most orders, legitimate researchers with names similar to those on the counter-terrorism list could be unfairly burdened by this screening system. Although a counter terrorism list would prevent known terrorists from ordering CSNA, it would do little to prevent hostile actors not on the list from acquiring the technology. As with previous customer screening methods, a counter-terrorism list would not improve biosafety since it would not prevent legitimate researchers from obtaining potentially dangerous CSNA that they are not trained to handle.

[255] Available at: http://www.wisconsinproject.org/risk.html
[256] Demonstration versions of the Risk Report are available at: http://www.riskreport.org/demojs/riskmain.asp

Licensing all CSNA Customers

The customer screening methods described above do not require any activity on the part of legitimate researchers. The next option for customer screening that we will present, licensing, requires action by all who wish to purchase CSNA. A licensing system could be modeled after the current system used to control the sale of radionuclides. This system would require customers wishing to purchase CSNA to apply for a license. The licensing agency would need to determine the best way to evaluate if a license should be granted which could involve an in-person visit from a federal or state regulatory agency (as is done for radionuclides). Because all researchers that currently work with CSNA (approximately 200,000 in the United States) would need to obtain a license, the amount of time it would take to implement this screening program, the cost associated with implementation, and the initial burden to researchers would be relatively high (options for implementing a licensing program are addressed in the section "implementation options for licensing"). We note, however, that the risk from misuse of radioactive materials was thought to be sufficient in the past to justify the cost.

Although the implementation costs associated with this screening program would be high, once customers received a license, screening could be performed by CSNA providers or by a government agency. We anticipate that post-licensing screening would be rapid and would not be expected to slow down the processing or shipping of orders. Implementation speed and costs could be cut by automatically granting licenses to those researchers already verified as being legitimate, for instance those that already hold a license to work with select agents (~8,000 individuals) or those that currently hold a license to purchase radioactive materials. Similarly, for large institutions, a senior license could be granted to the biosafety officer at that institution, who is in charge of granting sub-licenses to people at that institution. Compared with the customer screening methods described above, licensing would be more effective at improving sing biosecurity since hostile actors may find it difficult to obtain a license, especially if a face-to-face interview is required. This method is of course not infallible since it is possible that an individual wishing to cause harm may work for a legitimate company and therefore be able to obtain a license. Because, when used in the absence of sequence screening, licensed customers would be able to obtain any CSNA regardless of the sequence, licensing CSNA customers will not increase biosafety.

When considering international customers of US-based suppliers, it is unclear what licensing truly means. Does the US government invest licensing authority in some foreign institution or are US agents sent overseas for the purpose of licensing? Perhaps, if similar regimes are adopted by foreign governments, full recognition can be granted to licensees in other countries pending review of the foreign country's licensing process. Alternatively, existing, foreign customers of the CSNA industry could be grandfathered into licenses without any further investigation (see further discussion below). For collaborations between US select agent labs and foreign labs, the CDC is responsible for inspecting the overseas facility, but this happens infrequently at best.[257] Several other possibilities for the licensing of foreign customers exist (such as drafting a list of "legitimate" foreign institutions), but none is completely satisfactory because they involve fitting an operation that is primarily national to the international community.

Licensing Gradation

The previous option for customer screening requires licensing of all CSNA customers. Although this is the only licensing option available that does not require sequence screening, by incorporating sequence screening into the CSNA regulatory scheme, one can devise a system where licenses are required only to acquire CSNA associated with particular organisms. This screening system may or may not involve

[257]NIAID Select Agent Research Review and Approval Procedure for New and Continuing Grants that Include Foreign Institutions, July 24, 2006. Available at: http://www.niaid.nih.gov/ncn/grants/sagrantsproc.htm.

licensing of all CSNA customers, but would require those wishing to order potentially harmful sequences to hold a higher grade of license. The extra licensing requirement for those wishing to order potentially harmful sequences would further improve biosecurity and, unlike the customer screening methods discussed above, would help improve biosafety as potentially harmful sequences would only be available to those parties who had obtained safety training. While several licensing gradation options exist, one option that could further increase both biosafety and biosecurity would be the issuing of licenses for potentially dangerous sequences from only specific organisms. For instance, a researcher may be licensed to purchase sequences from one pathogen (such as *Yersinia pestis*) but not another (such as *Bacillus anthracis*). Alternatively, the system could issue one type of license for those who can work with any select agent or one type of license for those who can work with any pathogen. This system could build upon extant biosecurity measures, like the Select Agent Rule, in that anyone approved to receive select agents is automatically licensed to receive CSNA associated with those select agents. As discussed above, it is possible that this system may require only those who want to obtain CSNA associated with pathogens only, or that all CSNA customers must get a license and a higher grade license is granted to those who wish to work with pathogen-associated CSNA.

Because this system requires that CSNA sequences be screened, the speed at which a licensing gradation system could be implemented would be relatively slow (while the sequence screening system is being developed and tested) and the cost of implementation relatively high; however, the burden to researchers working with benign CSNA could be decreased if only those researchers wishing to obtain sequences from pathogens (which we estimate as less than 10 percent of all CSNA customers) were required to obtain a license. The next section discusses several options for the implementation of a licensing system.

Implementation Options for Licensing

With the introduction of new passport regulations, both US citizens wishing to travel outside of the country and personnel in the passport office have found that implementation of new requirements can be burdensome. Although fewer individuals would be required to obtain a CSNA license than those wishing to obtain a passport, if one chooses licensing as the mechanism that will be used to screen customers, consideration must be given for how such a system could be implemented. Below we discuss five licensing implementation options and weigh the burden to customers, CSNA providers, and the government against the biosecurity risks associated with each plan.

Full implementation after a short delay: When considering only the biosecurity issue, one of the best implementation strategies would be to require all customers to obtain a license in a very short period of time (30-90 days). At the end of the licensing period those without a license would be unable to purchase CSNA without first obtaining one. This system leaves only a small window of time for hostile actors to order potentially dangerous sequences ahead of the new regulations. Plus, given the rapidity with which molecular reagents could be designed, it remains questionable if any delay from passage of a regulation to its implementation could prevent a hostile actor from obtaining needed reagents in the interim. Although somewhat beneficial to biosecurity, this implementation strategy is riddled with problems for CSNA customers, providers and the licensing agency. With implementation taking place after only a short delay it would be difficult to notify and license an estimated 200,000 US customers prior to execution of the regulation. Many customers would likely find themselves unable to order CSNA once the requirement for licensing was put in place, leading to research delays for customers and lost sales for CSNA providers. Furthermore, the race to get large quantities of CSNA customers licensed in a short period of time would overly burden the licensing agency which would likely lead to long licensing delays. It should be noted that a cursory licensing protocol, designed to alleviate the burden of licensing all users immediately, would vitiate the security benefit of this scheme.

Full implementation after a long delay: The burden to CSNA customers, providers and the licensing agency could be eased by increasing the length of time before implementing a licensing requirement for CSNA purchase. Providing an additional year or two before implementing the licensing requirement would allow more opportunities to educate customers about the requirement and provide them with adequate time to complete the licensing process. It is likely that this increased delay would result in fewer unlicensed customers at the time of implementation, thereby protecting both research and sales. Given that the licensing agency would have more time to issue licenses, the burden on that agency would be reduced as well. This implementation program (similar to that used with the new passport regulations) may not fully lift the burden on any of these groups since many customers may postpone obtaining a license until just before the requirement is to take effect, thereby canceling out many of the benefits associated with this plan. Additionally this strategy has an increased security risk when compared to the shorter implementation plan because hostile actors would be given a longer period of time to obtain potentially hazardous CSNA. Those wishing to use the CSNA technology as a weapon would be given ample time to plan for their future needs and order sequences before the licensing rules took effect (although, in the absence of a full threat assessment, it is unclear how much time would be needed by a hostile actor starting from scratch).

Grandfathering existing customers: In an effort to keep the burden of licensing low while increasing biosecurity one could choose to phase in a licensing program. One option for phasing in a program would be to allow customers with an established history of ordering CSNA more time to obtain a license (or perhaps grant them a license automatically) while requiring that new customers obtain a license prior to their first CSNA purchase. This system would eliminate the possibility that existing customers would not be able to acquire CSNA, and, as the majority of CSNA ordered on any day originates from customers that have ordered before, the burden on the CSNA suppliers is also greatly reduced. If the licensing system is national (that is, there is a national "historical customer database" contributed to, and accessible, by all companies), existing customers could order with any company. If companies are unwilling to share customer data, customers may find it difficult to order from new companies. If it becomes more difficult for a customer to switch providers, a temporary decrease in competition could result, and possibly prices, synthesis speed, and quality could suffer. While the burden of lost orders for market leaders would be low because most of their business comes from existing customers, new and emerging CSNA providers would have little opportunity to gain new customers. Like the previous plan, this implementation system would ease the licensing agency's burden since grandfathered customers would have a longer period of time to apply for, and receive a license before the regulation would apply to them. This licensing system offers significant biosecurity benefits as it allows purchase of CSNA by only hostile actors who had to foresight to establish a customer history before the rules came into effect.

Proof of license application: A second option for phasing in a licensing system could require that all CSNA customers prove they have applied for a license prior to purchase. While this system would not prevent hostile actors from obtaining CSNA, it may discourage them from making such purchases because they must provide some data to a licensing agency. It would be made clear in the system that customers who obtained CSNA and were later denied a license would be investigated. Because customers would not need to wait until their license had been approved before purchasing CSNA, they could continue to order and receive CSNA. Although providers would need to verify that customers had applied for a license prior to shipping orders (a procedure that would likely involve the customer e-mailing or faxing proof of application) the financial burden to providers from lost orders would be minimal. Under this implementation strategy it is important that the licensing agency process applications quickly so that those that had ordered CSNA and were denied a license would have minimal time to use the product. Still, because not all CSNA customers order sequences on a regular basis, the burden on the licensing agency would spread over a longer period of time than if the short delay strategy was chosen.

Licensing required immediately for sequences from pathogens: A final option for the implementation of a licensing system is one that could only be used in conjunction with sequence screening. This system would require those ordering sequences derived from pathogens to immediately obtain a license to obtain their order. In contrast, those ordering sequences from non-pathogenic organisms do not need a license until some much later date. While this method would be burdensome to those customers ordering potentially dangerous sequences (or sequences falsely identified as being potentially dangerous), customers ordering sequences unassociated with pathogens would not be inconvenienced. CSNA providers may observe some economic consequences associated with this implementation program given that customers ordering sequences from pathogens may be temporarily unable to do so; however, these consequences would be significantly less than those stemming from the short delay implementation program because the system would affect only a small portion of their orders. Because licensing agencies would initially only need to register those customers that order sequences from pathogens, the number of individuals requiring licenses would be relatively low at first, thereby keeping the licensing agency's burden low. This implementation program would be effective at increasing biosecurity despite the relatively low burden to customers, providers and the licensing authority because potentially dangerous sequences could not be acquired without a license as soon as the system goes into force.

Table 6. Stoplight chart presenting the trade-offs associated with options for the implementation of a licensing regime.

The column in blue indicates a customer screening system that could only be used in conjunction with a sequence screen Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given parameter

Parameter	**Short Delay**	**Long Delay**			
		No Phase In	**Grand-fathered Licensing**	**Proof of License Application**	**License Required Immediately for Pathogen Sequences**
Burden to customers	🔴	🟢	🟡	🟡	🟡
Burden to market leaders	🔴	🟢	🟢	🟡	🟡
Burden to new or emerging companies	🔴	🟢	🔴 *	🟡	🟡
Burden on licensing agency	🔴	🟢	🟡	🟡	🟡
Biosecurity	🟢	🔴	🟢	🟡	🟢

** If the database of existing customers is shared nationally, the advantage of market leaders over new or emerging companies would be minimal (and the dot would be green).*

Verifying Identity under a Licensing Scheme

Once it is decided who should be licensed to receive what CSNA and how to issue licenses, a system must be devised to ensure that a supposedly licensed customer is who he claims to be. A very simple mechanism for verification could be that each licensee is issued an identifying number that must be entered with each order. The CSNA supplier then enters the license number and the customer name into a website that ensures that the license number and the customer name match. Slightly more complex, but perhaps more efficient methods, involve the creation of an "e-certificate" or a private/public key pair for each licensed user, as is common in e-commerce.

Table 7. Stoplight chart presenting the trade-offs associated with options for customer screening.
The column in blue indicates a customer screening system that could only be used in conjunction with a sequence screen Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given parameter

	Email/ Shipping Addresses	Credit History	Denied Parties List	Counter Terrorism List	Licensing All CSNA Customers	Licensing Gradations
Implementation speed	🟢	🟢	🟢	🔴	🔴	🔴
Cost	🟢	🟡	🟡	🔴	🔴	🟡
Effectiveness for biosecurity	🔴	🔴	🟡	🟡	🟢	🟢
Effectiveness for biosafety	🔴	🔴	🔴	🔴	🟡	🟢
Burden to researcher	🟢	🟢	🟢	🟡	🟡	🟡
Applicability to international customers	🟢	🟡	🟢	🟢	🔴	🔴

Sequence Screening Software

Any sensible regulation of the CSNA industry requires that customers be screened; however, a choice exists whether or not customer screening should be used in conjunction with sequence screening. Furthermore, those wishing to screen sequences must make choices about what screening software and what database of sequences of concern should be used. In this section we will discuss the sequence screening software options that exist and the costs and benefits associated with each option.

BlackWatch

One option for sequence screening software currently available to CSNA providers is Craic Computing's program, BlackWatch,[258] and a system could mandate that all screeners use this program. BlackWatch uses a BLAST based algorithm to compare CSNA sequences that are 20nt or longer to sequences from pathogens. Although the software does not require an exact match, program has a set of parameters that must be met before a sequence can be flagged as belonging to a pathogen.

This screening software could be used in a screening program run by the providers themselves or the government because all BlackWatch functions can be accessed via the web or downloaded onto company computers. The ability to use the program without sending sequence information outside of a company's firewall addresses many of the concerns of companies regarding the protection of proprietary data. Furthermore, search results are archived by the system and associated with the query sequence and customer number allowing for a complete audit of a customer's sequences.

The software is currently offered to CSNA providers free of charge and, according to the company's website,[259] "can be readily incorporated into automated workflow schemes," therefore we believe that a screening system using BlackWatch software could be implemented into a screening system relatively quickly. This hypothesis is reinforced by the fact that several synthetic gene companies we spoke with

[258] Although other sequence screening programs, such as OligoSleuth and pathoGENEDetective, work similarly, neither are ready for use at this time.
[259] http://www.craic.com/pdf_docs/BlackWatch_Datasheet.pdf

are already using the technology to screen sequences and, according to those providers, screening requires only minimal resources. However, because of the volume of orders filled by the oligo compared to the gene industry, and the fact that no oligo providers currently use BlackWatch to screen oligos, we envision that the adoption of BlackWatch by the oligo industry will not be as trivial.

Although the BlackWatch software is operational, in its current embodiment it may not catch all CSNA orders of concern. The software is designed to minimize false hits and may be too lax at catching slightly modified sequences. Although the program can be used to screen sequences as short as 20nt, it has been optimized for use with gene length sequences, and its effectiveness at screening shorter, primer length sequences is questionable, as shown in our review of screening tools, in Chapter 3, above. For example, with short sequences, the screening parameters can be overcome by adding six nucleotides (such as a restriction enzyme recognition site) to the end of a CSNA sequence or by changing one nucleic acid even if this change does not alter the protein encoded by the sequence. We recognize that the BlackWatch program was designed to be flexible and that the screening stringency could easily be adapted to the requirements of a screening system. In the absence of a risk assessment, it is impossible to determine if the stringency offered by BlackWatch is adequate or not. Furthermore, in any discussion of the further development needed to adapt CSNA screening tools to this system, BlackWatch should be considered a viable starting point.

Government-Set Parameters

A second option for screening software would be to simply mandate that CSNA providers operate a screening program that meets set standards, but may do so using any tool (BlackWatch, BLAST or their own algorithm). Because the exact requirements are unidentifiable in the absence of a risk assessment, it is difficult to predict the acquisition and operating costs CSNA providers would be faced with due to the amount of further tool development and false hit rate that would likely result. With this system in place, companies that have already developed their own screening software would be able to continue using it once updated to meet the government-set screening parameters.

Develop New Software

A final option for sequence screening software would be the mandated use of a new tool that has yet to be developed. This screening tool could be an improved version of BlackWatch (or one in which the parameters are adjusted), the use of software currently being developed (like pathoGENEDective) or the development completely new software. Although the development of new software would be expensive when compared to using currently available software and would require time to develop (therefore increasing the time before such a screening mechanism could be implemented), the software could be designed specifically around the parameters set forth by a thorough risk assessment and the needs of CSNA providers. New software could take advantage of established sequence alignment algorithms like BLAST or could develop a unique system for screening. Because the software would be developed with the federal guidelines and CSNA provider's needs in mind, the software could be designed to keep false hits, false misses, and CSNA providers' operating costs low.

To capture the biosecurity benefit of implementing oversight early, it may make sense to begin sequence screening with an existing tool, such as BlackWatch,[260] or set of tools that meets established standards, while supporting the development of improvements to screening tools that improve false hits and false misses rates while easily integrating into the processes of the CSNA industry.

[260] Before use of BlackWatch is mandated, it must be adjusted so that its screening parameters meet a false hit rate acceptable to the CSNA industry while possessing a false miss rate that is acceptable in the light of risk assessments.

Table 8. Stoplight chart presenting the trade-offs associated with options for screening software.
Green indicates a good choice, yellow a satisfactory choice and red a poor choice for a given category

	BlackWatch	Government Set Parameters	Develop New Software
False hit	🟡	VAR	🟢
False miss	🔴	VAR	🟢
Acquisition cost	🟢	VAR	🔴
Time to implementation	🟢	🟡	🔴
Operating cost For CSNA provider	🟢	VAR	🟢

Databases of Sequences of Concern

All sequence screening software requires the use of a database of sequences against which to compare orders. Several databases options exist for use with a sequence screening program. Below we will discuss the tradeoffs associated with options for database content and control.

Supplier Determines List and Database

One option for a sequence screening database would require CSNA suppliers to screen orders against potentially harmful sequences, leaving the determination of what sequences are potentially harmful to the CSNA providers, and mandate that each supplier develop and maintain their own database. This option would require few financial resources from the government but would be economically burdensome to CSNA providers. Compelling CSNA to update their databases when new, potentially dangerous sequences became known would be difficult as constant database updating would require additional financial resources. Furthermore, it would advantageous from a business standpoint for a company to develop a database with a minimum quantity of sequences given that CSNA orders matching sequences in the provider's database would cost the company time, money and potentially customers. In an effort to minimize false hits, false misses would likely be increased; therefore, this may not be the most effective way to prevent CSNA misuse. Moreover, due to a lack of consistent standards among CSNA suppliers, hostile actors may be able to obtain potentially dangerous sequences by identifying the company with the weakest database. Lastly, although the suppliers undoubtedly employ the best molecular biologists they are not necessarily familiar with data related to biological weaponeering and are likely to exclude important sequences and include irrelevant ones. It is important to recognize that although this option may seem palatable to industry wishing to avoid regulation, most companies we interviewed who perform screening wanted information on what to screen against and how, suggesting that this option may be unpopular because it does not provide enough guidance.

Nationally Mandated List

One option available for local screening would be for companies to screen sequences against a nationally mandated list. The relevant authority (for instance, an industry group or the federal government) would mandate a list of pathogens (for the full genome option, described below) or pathogenic sequences (for the partial genome option, also described below) that would need to be included in each company's database. The database itself would be created, updated and maintained locally by CSNA providers with guidance from the national authority. Alternatively, a third party could curate the database using the mandated list. This system would put most of the burden associated with database updating on the CSNA provider or third party database curator; however, using a list of sequences associated with pathogenicity (as opposed to the whole genomes of pathogens) would place slightly more burden on the national

authority as it would be their responsibility to determine what new sequences would be added and what new sequences should be removed as new discoveries are made. Because it is unlikely that new organisms would be added on a regular basis, the additional burden to the national authority in a full genome database system would be minimal. Because the burden associated with updating the database would fall primarily on the CSNA companies, compliance with database updating may be poor. This system is less secure than one run by the government because it allows those outside the government access to the information in sequence screening databases. This security vulnerability may provide hostile actors with an opportunity to learn what sequences are in a database, therefore providing them an opportunity to game the system.

Government (or Third Party) Database

One database option that could be used in either a government-run or local screening system would be a database curated by the government. If the screening is performed by the CSNA providers, the government would need to provide downloadable versions of the database that does not allow users to inspect its contents (a black box) in order to minimize risk that the system could be circumvented. The database could be updated automatically in much the same way computer virus software is currently updated, leading to a high level of compliance in regard to database updating. Unlike a government-mandated list, which would place most of the burden associated with updating the database on the CSNA providers, a government-run database would place this burden with the government. As before the cost associated with updating a database of sequences associated with pathogenicity would likely be greater than those associated with full genome sequences as the content of the database would likely need to be updated on a more regular basis.

Instead of the government itself, the sequence database could be maintained and updated by a third party. Responsible third parties could be those companies currently developing screening software (like Craic Computing, which already has a sequence database), a contractor, or a national laboratory. Responsible parties would have to be capable of maintaining database security.

Full vs. Partial Genome Sequencing

In this section, we discuss the tradeoffs of screening against the full genomes sequence from pathogens of concern or against only certain components from pathogenic organisms. Screening against full genome sequences may have a disadvantage in terms of false hits. For instance there is a high level of sequence homology between *Bacillus anthracis* and *Bacillus subtilis* (a benign soil bacterium that is used as a model organism in several labs, in fact, many taxonomic studies suggest that these two organisms are in fact the same species). Screening against the full genome sequence of *B. anthracis* would likely produce false positives for individuals working with *B. subtilus.* Also, from a purely statistical point of view, the more sequences one is screening against, the more chances one has of getting a hit. It should be noted that for certain viruses that have a small genome, their entire genome may be uncharacterized or known to be important for virulence, eliminating the ability to discriminate the sequences that may be excluded from the database.

This false hit rate could be reduced by carefully choosing the elements from the genomes of interest that will be included. This option may include screening against only known pathogenic components or only those components that are unique to the pathogenic organism. Alternatively, this option could involve excluding components known *not* to be associated with virulence of the organisms of interest (like the housekeeping genes of *F. tularensis*). It should be noted that while including only partial genome sequences would lead to a decrease in false hits it may also lead to an increase in false misses. Although it is unlikely that an adversary will identify a heretofore unknown pathogenic component before it is published in the literature, a pathogenic component may be flanked by two benign sequences. The

sequences, which would not be included in a partial genome sequence database could then be used to PCR amplify, if this is of concern, the pathogenic sequence in between.

Table 9. Stoplight chart presenting the trade-offs associated with options for sequence databases.
Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given category

Parameter	Supplier Determines List & Database	Full Genome Sequence		Partial Genome Sequences	
		Federally Mandated List	Federal Database	Federally Mandated List	Federal Database [261]
Cost to government for updating database	🟢	🟢	🟡	🟡	🔴
Cost to CSNA suppliers for updating database	🔴	🟡	🟢	🔴	🟢
Ability to game the system	🔴	🟡	🟢	🟡	🟢
Compliance with updates	🔴	🔴	🟢	🔴	🟢
False hits	🟢	🔴	🔴	🟡	🟡
False misses	🔴	🟢	🟢	🟡	🟡

The Microbes that could be Included in the Database

Two questions arise when attempting to determine which genomes should be included in the sequence database: which microbes should be included and which exemplars of these organisms should be included?

The first questioned is answered by deciding what activities by hostile actors the system should prevent. In this case, activities of concern can be divided into two categories: the synthesis (and manipulation) of viruses and the manipulation of bacteria. Clearly, when considering the synthesis and manipulation of viruses, it becomes clear that the database need only contain sequence information from viruses (and perhaps the few non-viral sequences associated with the increase in virulence of viruses, like the gene encoding interleukin-4) and while considering the manipulation of bacteria, the bacterial genomes must be included. Today, following established protocols, a hostile actor could use oligos to manufacture some viruses, obtaining an agent without any environmental or medical samples. In contrast, today an actor could only introduce a gene from a pathogen into a non-pathogenic bacteria, and through experimentation, eventually arrive at a strain that is more pathogenic (still, probably not as pathogenic as the strain that normally possesses the gene encoding the pathogenic element). As CSNA synthesis capabilities expand, it may be one day possible to synthesize the entire genome of a strain of bacteria and use it to replace the normal genome of a closely related non-pathogenic bacteria to get the desired strain. However, the challenges associated with this effort are daunting and several orders of magnitude more labor intensive than the synthesis of viruses. Furthermore, because bacterial genomes are, on average, two to three orders of magnitude bigger than viral genomes, and because bacterial genes are more similar to genes from non pathogenic microbes than viruses (especially closely related non-pathogenic bacteria), the consequence of including bacteria in the sequence database will be a vast increase in false hits.

[261] A federal database could be "black boxed" to maintain the secrecy of its contents.

It should be further noted that the activities of concern related to the manipulation of bacteria could be performed by direct synthesis of genes of interest or by PCR-based amplification of the genes from an environmental or medical sample. Given that custom PCR primers have been available for many years, this risk is not enabled by what can credibly be called an emerging technology. Given the relative ease of acquisition of some pathogenic bacteria from environmental samples, it remains an open question if the cost in false hits of the inclusion of bacterial sequence information in the database is worth the small mitigation of risk.

If bacterial sequences are to be included in the sequence database, it must be decided which species be included. Many process assessments have already been completed determining the relative ease of weaponization and degree of harm that would be caused by various pathogens, providing a good starting point for the determination of which bacterial genomes should be included. To determine which bacteria must be included in a CSNA screening system, however, requires combining information from these past process assessments with a risk assessment that determines the availability of species or strains closely related to the pathogen of concern and the ability to use sequence information from the pathogen of concern to change the related organism into something useful in a weapon.

To determine which viral sequences should be included in the sequence database, a similar process should be undertaken that considers the relative ease of weaponization and the degree of harm that could be caused by the virus, alongside the ability to synthesize the virus from oligos (or long double-stranded pieces of DNA or RNA). This assessment would consider the ease of synthesizing the genome in a useful state, the role of accessory factors in the early life-cycle of the virus and the ability of helper cells currently producing a related virus to produce useful, pathogenic virus. In addition to determining which viruses should be on the list, the assessment would indicate the point of control in viral preparation that would be prohibited. For example, poxviruses require non-structural proteins to be in the infectious virion in order for the viral DNA to be replicated in the newly infected cell. Therefore several points of control are evident for poxviruses, including the possession on the naked complete DNA sequence, the possession of poxvirus DNA and helper cells (the combination of which would allow for the production of viruse particles). Or possession of infectious particles.Alongside this assessment, one must identify the sequences needed to manipulate commonly available viruses (such as influenza virus) into agents that are truly dangerous without complete genome synthesis.

Once it is decided which species of bacteria and which viruses be included in the database, it must be decided which exemplars to include. This decision becomes important because the many genetic differences that distinguish two strains may cause screening systems to miss short oligos designed against a strain related to, but not identical to, the strains in the database. Beyond deciding which strain to include is the issue of deciding which exemplar of that strain. For instance, there are many sequences from *B. anthracis* Ames in the NCBI database, and some, for the same gene, have slight differences that could potentially lead to false negatives for short oligos. This problem is likely negligible if we are not concerned with the PCR-based amplification of a gene, as these small differences are unlikely to mask the detection of long sequences or *all* of the many small sequences required for the synthesis of an entire gene or viral genome. If PCR-based gene amplification is of concern, it may make sense to include in the database every sequence from all the strains of interest, because the sequences that are identical between the strains will not add to the false hit rate, and the differences will contribute to the reduction of the false miss rate. The ideal reference collection would faithfully represent all naturally occurring sequence diversity within each select agent taxon.

An Additional "Cost" of Sequence Screening

Beyond the time sequence screening requires, and beyond the cost from false hits (both in order delays and the cost for personnel to investigate the hits) there is an additional "soft" cost of sequence screening.

Currently, since no company screens oligo orders for content, these companies do not know what they are shipping to their customers. If a sequencing screening system were implemented, CSNA providers would gain knowledge about the content of their products. The most immediate implication of this knowledge is that, if Commerce Department Regulations are supposed to be applied to genes and/or oligos "associated with pathogenicity", shipping many CSNA orders overseas will require the CSNA provider either to knowingly ignore the regulations or to endure the untenably lengthy approval process. The former option may open CSNA providers to legal action; the latter will likely cause an unacceptable delay in order delivery that would cost the CSNA provider their foreign customers who wish to obtain sequences found in pathogens. If oversight of the CSNA industry is to be implemented, the Commerce Department should clarify its regulations to exactly define what the terms "genetic elements" and "associated with pathogenicity" mean so that CSNA providers know what they need to do to remain compliant with the law. Furthermore, if the regulations are redefined to cause genes and oligos that match sequences from listed pathogens to fall under regulation, the Commerce Department must streamline the approval process to enable foreign trade in CSNA to continue or must be willing to cause the loss of a significant fraction of the foreign customer base of US CSNA providers.

Even when shipping inside the US, knowledge of the sequence of the orders shipped may open CSNA providers to additional liability. Although thorough and conscientious customer screening could partially mitigate this risk, a CSNA provider that knowingly shipped genetic elements of a pathogen that is used in an attack may find itself the target of a lawsuit from a victims' group. We question if this risk is exacerbated by knowing the nature of the product because, in the example given above, a victims' group could claim that customer screening methods were not stringent enough and that screening sequences to determine if the product is potentially dangerous, although not standard industry practice, is certainly possible.

Conclusions

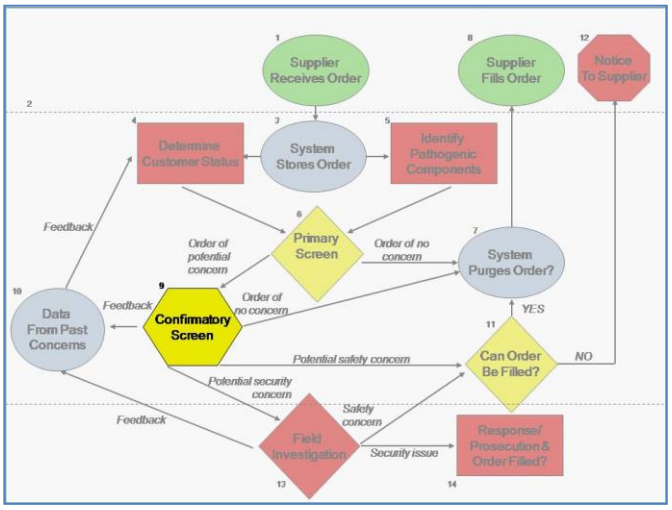
When choosing whether or not to screen CSNA orders, it is important to consider the tradeoffs of each oversight option. Customer screening alone can be an efficient mechanism that requires minimal operating costs. Although, depending on the method used to screen, its usefulness in improving biosecurity varies, customer screening alone is not an effective regulatory mechanism for boosting biosafety because it does not include a mechanism to prevent legitimate researchers from acquiring potentially dangerous CSNA that they are not trained to handle. The addition of sequence screening may improve biosafety by allowing only authorized researchers to order potentially hazardous CSNA sequences, thereby preventing accidents that may occur when such sequences are obtained by well meaning, untrained persons. Furthermore, sequence and customer screening may be more effective at improving biosecurity than customer screening alone. Hostile actors who are able to avoid detection by a customer screen will could be permitted to order only innocuous CSNA, as the purchase of sequences identified as potentially dangerous would require further authorization as in the case of licensing.

Table 10. Stoplight chart presenting the trade-offs associated with options for customer and sequence screening versus customer screening alone.

Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given category

	Screening Customer	Screening Customer and Sequence
Operation costs/efficiency	🟢	🔴
Biosafety	🔴	🟢
Biosecurity	🟡	🟢

Step 9: Confirmatory Screening



When primary screening, which considers customer information possibly alongside sequence information, results in the generation of a hit, the screener may take additional actions before passing on information about the hit to a field investigation and making the decision on whether or not to fill the order. It should be noted that one option is to not have confirmatory screening and to simply pass on all hits from primary screening for field investigation, or, more commonly, simply rejecting any order that fails the first screen. We assume that even in a well designed system, the majority of hits from a preliminary screen would be generated by well-meaning individuals and therefore should not be passed on to field investigations or rejected outright just for expediency. In this section, we discuss options to resolve hits generated by a preliminary screen to reduce the quantity of legitimate orders denied and reduce the burden on field investigation teams.

Non-interactive Screening

This option for secondary screening does not involve any contact between the screener and the customer. Examples of activities (many which could be automated) under this description are:

- If an order is received from an unfamiliar or non-official domain name (such as Gmail, Yahoo! or Hotmail), the screener may verify the customer's identity by searching the website or the directory of the customer's home institution.

- If the order is from an individual whose identity cannot be verified, the screener can investigate the sequence requested (assuming this is not done as part of a primary screen). If the customer is requesting a sequence that seems harmless, the order can be filled.

- If the order produced a hit because it matched a sequence that the customer is not allowed to obtain, the screener can investigate other orders made by the same customer to see if they have produced hits. This method would greatly reduce hits caused by serendipitous matches between innocuous organisms and pathogens. If the customer placed orders for many oligos and is from a known, legitimate institution, the order possibly should be filled. We note that this method is most useful if a screener has access to orders system-wide; if not, an adversary wishing to circumvent the screen could place one order (and a few innocuous orders to screen the order of interest) with a few companies to obtain the desired reagents.

- As an alternative to the method above, other orders from the same customer could be screened for content. If the sequences of the other orders match a single organism that the suspect order also matches (in addition to the pathogen) it is probably a serendipitous match and possibly should be filled if the customer is from a known, legitimate institution.

Phone Call to Biosafety Officer

Alternatively, if an order produces a hit in primary screening, a call to the institutional biosafety officer may help address questions of whether the order is going to a person/facility with approval to work on the organism ordered or not. Alternatively, the biosafety officer may issue an assurance that the researcher will work only with molecular components of the pathogen that are non-hazardous in isolation. This method is only likely to be useful for biosecurity purposes at larger institutions where the biosafety officer is not the same person placing the order.

Phone Call to PI

This type of activity may be the most common if customer screening occurs alongside sequence screening because an oligo ordered may match the sequence of a harmless organism and a pathogen by chance. Therefore, a phone call to the PI would be necessary to determine what the target of the oligo really is, and, if the PI mentions an application consistent with the sequence, perhaps the order should be filled. This phone call may be perceived as more costly than calling the biosafety officer, as a biosafety officer should be expecting these types of calls, whereas a PI may not. Furthermore, a PI may not be immediately available to field a call, leading to delays in filling the order.

Issues Regarding New Biotechnology Companies and Amateur Biotechnologists

When discussing means of secondary screening to resolve hits produced in primary screening, we find that there are very few efficient means of determining that a new biotechnology company is legitimate without a visit to the company. An adversary could easily establish a company with a good credit record, history of interaction with equipment and reagent suppliers and a realistic website. Calling the biosafety officer of such a company may simply be asking one hostile actor to verify the identity of another. For this reason, we estimate that orders from new biotechnology companies are going to be very difficult to resolve to the satisfaction of both the security and the industry communities.

Some possible solutions involve the screener calling the biosafety officer of the new company and obtaining references from other scientists in the field or from their financial supporters. If their financial supporters are established venture capital companies, then the customer can be viewed as legitimate, because of the resources spent by these organizations to vet the financial promise of their portfolio. Similarly, if the references given by this company are from scientists in the field who would be allowed to order similar reagents, the order should be filled if the references can be verified.

Importantly, the relative value of each method of secondary screening, and the concept of operations that would guide the decision of which step is attempted first, greatly depends on the nature of the hit in the primary screen and the methods used that generated the hit.

Let us take the following example of a sensible concept of operations. Assume that the primary screen involves checking the order for content related to select agents and an order is flagged because it contains sequence from a select agent and is ordered by a person who is not approved to work with oligos from select agents. In this case, the first move may be to call the PI of the lab that ordered the oligo and ask them what it was for. If they claim that they are working on *Arabidopsis* (a plant commonly used in molecular studies) and the sequence matches a select agent AND *Arabidopsis*, the order should be filled. If the PI is not available, the screener could see if the customer ordered other orders that were flagged, or ordered many oligos, all of which match from *Arabidopsis.* If any of these conditions are fulfilled and the researcher is from a known, legitimate institution, the order should be filled.

Another group that may pose problems for oversight in the future is composed of the amateur biotechnologist not affiliated with an institution. Similarly to amateur computer programmers, the reasons for amateurs to be drawn into the manipulation of creation of life using the tools of molecular biology are likely vast and unfathomable. Likewise, it is difficult to imagine oversight regimes that could ensure that these amateurs will handle CSNA safely and responsibly.

The risks of granting amateur biotechnologists, potentially working with no safety equipment and little training, access to CSNA are difficult to quantify but easy to characterize. Although most will not be capable of producing anything dangerous due to lack of experience, an unsupervised and untrained amateur has the potential to harm himself or others. Furthermore, any oversight regime that allows amateurs to gain access to CSNA from well-known pathogens will probably be perceived by the public as dysfunctional.

Perhaps the easiest solution to the "problem" of amateur biotechnologists is to prevent them from acquiring CSNA, thereby reducing the risk that the technology could be misused. However, this solution has real costs. Preventing this group from acquiring needed CSNA reagents may stifle an important driver of innovation in biotechnology (if the analogy to computer programming holds). Furthermore, if amateur biotechnologists become numerous, they may represent a lucrative market that cannot be served by American companies. Building on this point, excluding amateur biotechnologists may produce a black market for CSNA that those with less noble intentions could exploit.

Another solution is to screen CSNA orders so that customers unaffiliated with a recognizable institution cannot obtain any CSNA associated with pathogens. If screening all CSNA orders is considered too burdensome, perhaps only orders from amateurs could be screened to ensure that the CSNA stands is unlikely to be misused (or perceived as dangerous by the public). As long as amateur biotechnologists are rare, these steps will not be overly burdensome.

One possibility that would allow amateur biotechnologists not affiliated with an institution to legally purchase and safely use these tools would be a regime of training and licensing program similar to what exists in some states for potentially dangerous pesticides. The Iowa Department of Land Stewardship's Pesticide Bureau requires licensing for both the sale and purchase of peptides. Individuals wishing to purchase or apply pesticides are required to obtain a license to do so. Licenses are acquired and renewed by passing an examination or completing an instructional course. [262] Like Iowa, many states require such training and certification to ensure these chemicals are handled safely and properly.[263] Using a certification program such as the one described above as a guideline, a certification process focused on synthetic nucleic acid biosafety could be developed for individuals wishing to obtain CSNA. Such a program could help prevent issues associated with accidental misuse of CSNA as well as prevent the establishment of a black market.

---

[262]Iowa Department of Agriculture Pesticide Bureau. Laws and Rules Administered by the Pesticide Bureau. Chapter 206 Pesticides. http://www.agriculture.state.ia.us/pestlaws.htm#206.5%20Certification%20requirements--rules

[263]Washington Department of Agriculture. Pesticide/SPI Licensing. Oct. 31, 2007. http://agr.wa.gov/PESTFERT/LicensingEd/Licensing.htm#GettingLicensed, Oregon Department of Agriculture. Pesticides Division. Pesticide testing and licensing. http://www.oregon.gov/ODA/PEST/licensing_index.shtml//prohort.ifas.ufl.edu/Pesticide.html

Step 10: Order Storage

A CSNA screening system requires the review of customer and, in some cases, sequence information. From our industry interviews we determined that nearly all companies currently archive both customer and sequence information for two or more years (although a few companies mentioned that the data was not stored in a searchable format). In the event CSNA technology is used in a biological attack, past order information may be useful in identifying the individual or individuals responsible for the attack. Importantly, in this section we discuss how information will be stored for retrieval in unusual circumstances (such as a forensic investigation). In the section describing who will implement the oversight, above, we describe how data from screening can constantly be used to improve primary and confirmatory screening.

Below we outline several options for archiving CSNA order information, ranging from allowing providers to each maintain their own database to storing customer and sequence information from all CSNA orders in a central, government-run database. The information in this section is most relevant if companies screen their own orders, because we describe here the costs and benefits of sharing data with a central database.

No Data Leaves Company

One option for storing CSNA order information is to require providers to archive all order information for a given period of time in an onsite database. Given that all CSNA providers we spoke with currently achieve this information for at least two years, this option would require almost no effort to implement, except for that required to re-organize some databases into a searchable format. Used in conjunction with a local screening system or a government-run system that reviews only customer (and not sequence) information, this data storage option preserves a company's ability to keep customers' CSNA order information confidential. In the event CSNA technology is used in a biological attack (or the attack is prevented), each supplier's database could be searched for information that could help lead to the identification and conviction of the group responsible for the attack. With 40 or more company databases to search, each likely using a different storage system, sorting through past orders may be difficult and time consuming. Because data storage has become less and less expensive over time, the only real advantage of this archiving system is that it affords the ability to maintain the confidentiality of customers' CSNA orders and would therefore be palatable to industry.

Central Storage of Confirmatory Screen Data Only

A second option for storing CSNA order information is to deposit customer and sequence information from orders that do not pass the primary customer and/or sequence screen into a central database. This option, when used with a local screening system or a government-run customer screening system would protect the confidentiality of most CSNA orders, as only those orders called in to question would need to be transmitted offsite for storage in a national database. In the event of misuse of CSNA, any CSNA order that had been called into question could be searched quickly, hastening a forensic investigation. On the downside, any order not flagged in a primary screen would not be centrally archived (although this information would still likely be stored by the provider). Therefore, if the CSNA customer responsible for the attack had escaped any kind of detection by the screening system, this option would not provide an advantage over the previously described storage option.

Certain Orders Stored Centrally

The next option for archiving order information could only be used in combination with a CSNA sequence screening program. Orders for potentially hazardous sequences would be stored in a centralized

federal database. This option would protect the confidentiality of most customers, as only those ordering sequences appearing to be from certain pathogens would have their order information transmitted offsite for storage. Since it is unlikely that sequences from non-pathogenic organisms would be used in a biological attack if such an event were to occur, this option would allow for rapid searching of those orders of greatest interest. Once again, if the screening system did not recognize the sequence as being potentially dangerous, the order information would not be centrally stored.

Table 11. Stoplight chart presenting the trade-offs associated with options for storage of order screening data.

The column in blue indicates a customer screening system that could only be used in conjunction with a sequence screen Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given category

	No Data Leaves Company	Info From Confirmatory Screen	Certain Orders Stored Centrally	Only Store Denied Orders	All Data Centrally Stored
Holding confidential info	🟢	🟡	🟡	🟢	🔴
Forensics/Attribution	🔴	🟡	🟡	🟡	🟢

Central Storage of Denied Orders Only

Although all of the companies we spoke with currently archive past order information, it is unclear if they kept records on orders they decided not to fill. For this reason it may be advantageous to introduce a central database for storage of denied orders. Because CSNA may be self-synthesized or purchased from a foreign CSNA providers, refusing a CSNA order will not necessarily prevent an individual from obtaining a desired product. In the event of a biological attack (or other forensic investigation), it may be important to review denied CSNA orders for potential leads about who might be responsible. This archiving method would prevent the transmission of confidential sequence information for all legitimate orders and therefore should be acceptable to industry. While storing denied orders is important, it would only be useful if hostile actors were not successful in ordering CSNA from a US provider. Information on orders that evaded the screening system altogether would only be stored by the CSNA providers' normal system.

All Order Information Centrally Stored

A final option for archiving CSNA orders is to store all orders in an offsite, federally run database. In this system, all orders, including information about the customers name the sequence ordered would be stored offsite. While this option could be used in combination with a local screening system, the benefits associated with increased customer confidentiality in a locally run system would be lost. This option for order archiving would provide maximum forensic benefits in the event of a biological attack or investigation since all orders could be easily searched simultaneously.

Step 13: Field Investigation

Because law enforcement investigations in this area are by their nature sensitive, it would be difficult to speculate on the precise roles taken by law enforcement agencies when investigating someone wishing to misuse CSNA. Nonetheless, we will attempt to provide the broad outlines of how a field investigation might take shape. The following discussion of options is heavily informed by ongoing experience with the Select Agent Program administered by the CDC and APHIS and enforced by the FBI. We could think

of no other options that made sense for the field investigation of CSNA orders, so we simply describe how a field investigation would operate.

Under the Select Agent Program, CDC and APHIS inspectors are given the authority to inspect the facilities and records of any registered entities. In the event that suspicion of a safety or security violation arises, whether in course of inspections, or by some other means (such as a tip), further in-depth investigations are authorized by the regulatory authority (either the HHS Secretary in the case of CDC-registered entities or the Secretary of Agriculture in the case of APHIS-registered entities). Further investigations may also be conducted by the Inspectors General of HHS and USDA, who are authorized to pursue civil penalties against any entity or individual found to be in violation of the Select Agent Rules. In the event that a suspicion of criminal misconduct arises on the part of CDC/APHIS or the Inspector s General of HHS and USDA, the case will be referred to relevant officials within the FBI WMD Operations Unit at FBI Headquarters through a well-established and robust liaison system.

Given that existing regulations are vague when CSNA is concerned, one could imagine that the system uncovers conduct that is not criminal (because current criminal regulations do not cover owning fragments of the smallpox, virus, for instance), but certainly worrisome. Certainly, most activity of this type would still fall under the purview of the FBI because it would be assumed that the intent of actor was to eventually cause harm using biologicals. Although past history of incidents with biological agents teaches us that we should also expect customers with intents that are hard to fathom. One could imagine individuals who collect genomic fragments of all the select agents "just to have them" or who wish to make art incorporating the DNA from these agents or an individual who claims, like Larry Wayne Harris did, to be privately researching a cure to protect the US from "super-germ-carrying rats".[264] It should be noted that incidents such as the latter were a major component behind the momentum to clarify the law and draft the Select Agent Rules.

The FBI is also in the process of establishing lines of contact between CSNA providers and FBI WMD coordinators located in local field offices. Information received by WMD coordinators in the field would then be fed back to the WMD Operations Unit. However, this raises the question of what the company is expected to do with a suspect order. If the order is filled, the company may expose itself to liability for any safety or security consequences resulting from misuse of the order. However, if the company refuses the order, it may have the effect of alerting any would-be criminals or terrorists, thus compromising the ability of law enforcement to apprehend them.

Concept of Operations for Suspect Orders

Given the large volume of CSNA orders received daily, it is likely that companies will receive questionable orders on a semi-regular basis. For this reason a Concept of Operations must be defined by the federal government to guide providers when such an event occurs.

Depending on the type of primary screening performed, an order could fail confirmatory screening for biosafety or biosecurity reasons. If the order fails for biosafety reasons, a dialog should be established between the supplier and the biosafety officer of the customer to resolve the issue (or to explain why the order should not be filled). If the biosafety officer *is* the customer (as in the case of small biotechnology companies), or if, during a conversation with the biosafety officer it is clear that the customer will try to obtain the suspect reagents in the future, the CDC or USDA should be contacted. Even if no nationwide screening system is adopted (or prior to the adoption of a system), the CDC and USDA should provide points of contact to CSNA providers in case of suspect biosafety issue.

---

[264] Quote from Larry Wayne Harris found in Harrison, L. "Larry Wayne Harris", Encyclopedia of Bioterrorism Defense, Pilch and Zilinskas eds.; Wiley-Liss, Inc, Hoboken, NJ, 2005.

If the order fails the confirmatory screen because there is suspicion about the intent of the customer (a biosecurity concern), the CSNA provider should contact the WMD coordinator at the FBI's field office, or possibly, a representative from DHS's Homeland Infrastructure Threat and Risk Assessment Center (HITRAC). Although the FBI currently is the body that investigates allegations of the intentional misuse of biological materials, the suggestion that HITRAC should be involved is compelling because this center was established for the sharing of intelligence and risk information between DHS and private industry and an investigation by the FBI may be unwarranted if the suspicious act is not necessarily criminal (like the synthesis of parts of the genome of a pathogen).[265] However, it should be noted that the biotechnology industry, outside the component that provides pharmaceuticals and vaccines, is not listed as a critical infrastructure and the description of the chemical industry is clearly not meant to cover companies that manufacture CSNA. No matter who performs the investigation, currently CSNA providers generally do not know who to call, and therefore, even if no nationwide screening system is adopted (or prior to the adoption of a system) each CSNA provider should be given the contact information for the WMD coordinator in their local FBI field office or the appropriate contact at HITRAC.

After reporting the order that is suspect for biosecurity reasons, CSNA providers must be given guidance on what to do with the order. Two options immediately present themselves: the so called red light or green light options.

Under a red light system, providers are instructed not to fill orders that fail a confirmatory screen for biosecurity reasons unless specifically directed to do otherwise by the FBI WMD coordinator or HITRAC point-of-contact. If a red light system is adopted, the CSNA provider must be given guidance as to what to tell the denied customer. Possible responses could include: they are told not to communicate further with the customer, they are told to notify the customer that his order can not be filled due to a biosafety or biosecurity concern, they are told to notify the customer that his order can not be filled without specifying a reason, they are told to provide the customer with an alternative reason, that is, lie to the customer. Although the communication guidance given to the CSNA provider is likely to vary depending on the details of the refused order, communication guidance should be given every time a suspect order is reported. Otherwise, in the event the suspect order is from a legitimate customer, failure to fill the order without a reason given is likely to reflect very badly on the company, and, through word of mouth, may lose them more than just the customer that was flagged. One advantage of the red light system is that it offers law enforcement flexibility in determining how best to proceed with an investigation. Depending on the status of an investigation, law enforcement officials may prefer to withhold potentially enabling sequences. In the event they feel it is advantageous to release the sequences providers can be instructed to do so.

Under a green light system, CSNA providers will fill all orders unless instructed otherwise. Implementing the green light system would prove less burdensome to CSNA providers as they would not be required to hold questionable orders by default. Additionally, this system avoids the problem of alerting hostile actors of an investigation into their activities and problem of deciding what to tell a denied customer.

Though the proceeding text has dealt exclusively with orders from domestic customers, a similar concept of operations should be developed for the investigation of orders from foreign customers of US companies. Consideration should be given to which agencies (whether FBI, HITRAC, Interpol, CIA or some combination of them) would be best-equipped to conduct such an investigation.

---

[265] Chaparro, J. "Briefing to the Intelligence, Information Sharing and Terrorism Risk Assessment Subcommittee of the Committee on Homeland Security, United States House of Representatives", July 26, 2007. Available at: http://homeland.house.gov/SiteDocuments/20070726123048-73956.pdf

Supporting Regulations

Systems for the oversight of the CSNA industry could be underpinned by a variety of mechanisms: Industry Self-Regulation, Voluntary Federal Standards, Incentivized Federal Standards, and Mandatory Federal Regulations. In evaluating the costs and benefits of these various regulatory models that have been proposed for the CSNA industry, we consider five separate metrics: the likely need for government rule-making (at the agency, departmental, or legislative level), the likely palatability of the proposed regulations to the CSNA Industry and its customers in the research, biotechnology, and drug discovery communities, the ease with which the regulations could be internationalized, the likely uniformity of the standards, and finally, the likely degree of industry compliance with the standards. The overall advantages or disadvantages of each regulatory option are however, difficult to assess and are likely to be heavily influenced by decision-makers' priorities.

Industry Self-Regulation

Under this model, CSNA providers will, individually, or in groups such as the ICPS (International Consortium of Polynucleotide Synthesizers) voluntarily adopt a set of safety and security standards. That is, no rules, regulations or standards will be adopted industry-wide. Such an approach to regulation avoids the sometimes contentious and protracted process of government rule-making and is likely to prove most palatable to the CSNA industry and its customers. It is, after all, difficult to imagine an industry group adopting extremely onerous standards to which most of its members strenuously object. Proponents of self-regulation proposals (such as those offered by Maurer et al.) thus argue that it will be less likely to impede continued scientific advances and commercial growth[266] and point to the Asilomar protocols established in 1975 as a demonstration "that biological research communities can and do adhere to voluntary standards."[267]

However, the Asilomar model does little to address biosecurity challenges. The discussions at Asilomar leading up to the establishment of the protocols skirted around the issue of biosecurity and never confronted the threat of bioterrorism, which, at the time, was not a well recognized threat. Self-regulation, in short, offers only limited means with which to ensure compliance with a particular set of standards and offers no guarantee of uniform standards. Reliance upon a system of self-regulated sequence screening for example, may create an unintended market space for non-compliant CSNA providers offering unscreened sequences "no questions asked." Although the market for those who wish to order CSNA without oversight is likely very small and not enough to support a company, if screening procedures are costly, these companies could offer CSNA products more cheaply than their rivals and thusly gain market share. Some proponents of self-regulation, such as Maurer et al. also argue that self-governance schemes are intrinsically international in nature. However, as previously discussed in Chapter 2, this argument proves both empirically and logically unconvincing. Certainly, one cannot imagine why self-governance should be *more* intrinsically international than any other regulatory model and regulations adopted by the US have served as the model for regulations in many countries.

Voluntary Federal Standards

This regulatory model draws upon the approach taken by the Federal Consumer Product Safety Commission (CPSC) in developing safety standards. By statute, the CPSC can only issue a mandatory safety standard when there is no existing voluntary standard in place or if the industry fails, demonstrably to abide by those standards.[268] Thus, the CPSC has had extensive experience in working with industry to

---

[266] Bügl et al., 2006.
[267] Maurer et al., 2006.
[268] US Consumer Products Safety Commission, Regulatory Reform Initiative, Summary Report. June 1995. Available at: http://www.cpsc.gov/businfo/8005.html

develop voluntary standards. Industry is motivated to comply with these standards both by the desire to avoid the imposition of mandatory standards and by the wish to reduce liability in the case of accidents involving their products. While adherence to voluntary federal standards by consumer products manufacturers is no guarantee against tort claims, it is certainly a powerful legal argument that can reduce a company's legal vulnerability.

While the establishment of such voluntary standards may not require regulatory agencies to go through a formal regulatory process, let alone a legislative one, it nonetheless requires that they develop these standards through a consultative process with industry and other stakeholders. It thus requires a greater extent of government involvement in rule-making than the self-regulation model. Nonetheless, given the consensual nature of such voluntary standards, as well as industry's likely intimate involvement in their development, CSNA providers are unlikely to object to their implementation. Furthermore, as discussed in Chapter 3, many companies seek guidance from the federal government regarding screening practices.

Voluntary standards implemented within the US CSNA industry may well spread through an industry-wide process of diffusion of best practices and standards, but there are certainly no apparent reasons to think that they will spread internationally any more quickly than standards adopted under alternative regulatory models. However, in contrast to industry self-regulation, the collaborative involvement of industry and regulatory agencies is likely to result in a uniform set of safety and security standards across the CSNA industry. Nonetheless, this regulatory model raises similar concerns as self-regulation with regard to compliance. While companies may be better able to quantify the costs and benefits of compliance (in terms of vulnerability to litigation), it is possible that at least a portion of companies will choose not to comply if the costs for adopting a screening system is high.

Incentivized Federal Standards ("NIH Guidelines")

Several commentators, including the NSABB Working Group on Synthetic Genomics have recommended that the US government "require federal grantees and contractors to order from providers that screen and retain information about requests for Select Agent sequences following standards and practices developed by relevant federal agencies."[269] This, and similar proposals, are inspired in large part by current regulations found in the NIH Guidelines for Research Involving Recombinant DNA Molecules. Under the guidelines, discussed further in Chapter 4, individual researchers and institutions receiving NIH funds are required to abide by certain safety and containment standards and to undergo specified approval procedures depending upon the nature of their experiments. Based on the Asilomar protocols agreed upon in 1975, the NIH Guidelines have become widely accepted by the research community and biotechnology industry where a broad consensus has emerged that the Guidelines have "placed few obstacles in the path of scientific progress."[270]

The way in which Industry and researchers would regard the application of this approach to the regulation of CSNA providers and their customers would likely depend upon the breadth of the regulations. While under the NIH Guidelines all researchers associated with institutions receiving *any* NIH funding for work involving recombinant DNA are required to abide by its provisions, other NIH regulations (such as the limitations placed upon the use of federal funds for embryonic stem cell research) apply only to individual research groups or laboratories rather than to institutions as a whole. Industry is thus likely to object much more strenuously to the implementation of the former system than to the latter, as it likely captures a much larger proportion of their customer base. As a result, non-compliant companies are likely to face powerful market pressures to come into compliance with NIH-mandated standards.

[269] NSABB Working Group on Synthetic Genomics, "Addressing Biosecurity Concerns Related to the Synthesis of Select Agents", December 2006.
[270] Tucker, 2006.

Therein lies perhaps the greatest advantage of regulations modeled after the NIH Guidelines: the relative ease with which they can be internationalized. Since the US market is likely to remain the largest and most lucrative component of the global CSNA market for the foreseeable future, international CSNA providers will find themselves under the similar pressure to comply with the regulations as their US-based competitors. Likewise, US companies would not be able to avoid adopting NIH-mandated safety and security measures by off-shoring their businesses, unless they were willing to neglect the overwhelming majority of potential US orders. This regulatory system would have the further advantage (shared with all government-promulgated systems) of regulatory uniformity.

However, the NIH Guidelines place no obligation upon private companies and institutions that do not receive US government funding, entities that are certainly numerous enough to constitute a large and lucrative market for non-compliant CSNA providers within the US and in other countries. Moreover, while as previously noted, the NIH Guidelines, to a large extent, depend upon Institutional Biosafety Committees (IBCs) to provide oversight of potentially hazardous experiments, many institutions, especially small biotechnology companies, in practice, lack sitting IBCs. Additionally, the oversight record of IBCs has been, at best, mixed, a situation that is unlikely to improve with the addition of biosecurity concerns to their responsibilities.

Mandatory Federal Regulations (New or Redefined)

The relative shortcomings of fully or partially-voluntary regulatory models with regard to biosecurity concerns have lead some to propose more extensive government regulation of the CSNA industry. These proposals range from plans to simply redefine and update existing regulations to accommodate technological changes, to the promulgation of new regulations, or even new legislation. In the first case, as noted by Garfinkel et al in the Sloan Foundation Report on Synthetic Genomics, "even if policymakers believe that the current regulatory approach can accommodate synthetic genomics, guidance is needed merely to apply existing regulation or governance mechanisms, defined in the context of one set of technologies to a subsequent technological generation."[271] In particular, as argued by the NSABB Working Group on Synthetic Genomics, additional regulation of the CSNA industry may become necessary simply to maintain the current protections offered by the Select Agent Rules (which would be rendered irrelevant were malefactors able simply to synthesize or assemble synthesized sequences of controlled agents or toxins).[272] Thus, updating current regulations might take the form of the re-interpretation of the Select Agent Rules such that possession of pieces of CSNA, when put together, comprise the complete genome (or, some defined portion thereof) of any Select Agent or gene for a Toxin is controlled under the Rules. Alternatively, new regulations might define a broader or narrower group of sequences of concern which could be controlled via the Select Agent Rules or completely new regulations governing sequence screening and/or customer licensing (for instance, establishing less burdensome procedures for licensing and safely working with CSNA associated with select agents than the current Select Agent Rule provides for working with pathogens). Policymakers might also consider the NSABB-recommended harmonization of the CDC/APHIS (Select Agent) and Export Administration Regulations (Commerce Control List) lists of controlled agents and toxins. Further, some regulations, such as those that control the export of genomic material, may already possess the language necessary to regulate the industry, and simply need to be clarified or enforced. Any or all of these regulatory options may require extensive government rule-making, up to and including new legislation.

While Maurer et al. and others have argued that government-mandated regulations, due to their intrinsically national nature, are a poor way to approach the governance of the global CSNA market, these arguments are in our view, as previously discussed, only superficially convincing. One need only

[271] Garfinkel et al., 2006.
[272] NSABB Working Group on Synthetic Genomics, 2006.

examine the adoption of Select Agent Rule-inspired biosafety and security regulations in Britain and other European countries (EU-wide regulations are currently being considered) or the *intrinsically international* export protocols adopted by Australia Group nations to demonstrate that there is certainly no prima facie reason to believe that government-mandated regulations are more difficult to internationalize than industry-adopted ones. It should be noted that, if similar legislation is not adopted internationally, companies could escape regulation by relocating production facilities overseas (if regulations covered both suppliers and customers, this possibility would be avoided).

Naturally, one would expect that government-mandated regulations would be uniform in nature and that compliance could be ensured through administrative, civil, and criminal penalties as well as a potential government inspection regime analogous to that employed under the Select Agent Rules or other regulations.

Table 12. Stoplight chart showing trade-offs for regulatory schemes underpinning the oversight system against several parameters.
Green indicates a good choice, yellow a satisfactory choice, and red a poor choice for a given category

	Industry Self-Regulation	Voluntary Federal Standards	NIH-like Guidelines	Mandatory Federal Regulations (New or Redefined)
Need for government rule-making	🟢	🟡	🟡	🔴
Palatability to industry and customers	🟢	🟢	🟡	🔴
Ease of internationalization	🟡	🟡	🟢	🟡
Uniformity of standards	🔴	🟢	🟢	🟢
Compliance with standards	🔴	🔴	🟡	🟢

Cost of the Notional System

In this report, we present options for each component of a notional CSNA oversight system. Unfortunately, because many options are presented and each option has its own funding requirements, it is impossible to derive a single estimate of the total cost of an "average" notional system. Once options are narrowed by policymakers, we could present a manageable set of costs for candidate systems. In this section, we describe some concrete cost elements to provide information to policymakers when making decisions about the options presented above.

Customer Screening

Above, we presented several options for customer screening. Many, such as credit checks and the validation of e-mail and shipping address, are already performed by many CSNA providers. As these steps are already taken by most suppliers, the additional cost burden for companies that do not currently perform this type of customer screening is obviously absorbable into the cost structure of a CSNA provider.

Customer screening by comparing the customer's name to those on Denied Parties Lists or counter-terrorism lists would require CSNA providers to perform activities that they do not normally undertake. As discussed in Chapter 4, various Denied Parties lists are free, other databases require a user fee ($2,800/year). A counter-terrorism list would have to be established and maintained by the government to

prevent its contents form becoming public. As Denied Parties lists are already compiled, and software exists to screen customers against these lists, no development costs are required. The costs to develop a counter-terrorism list is more extensive, and, insofar as it does not overlap with the No-Fly List, requires intelligence collection and analysis to compile, the cost of which are difficult to quantify but would certainly dwarf the costs to build the interface and maintain a server. Clearly, the cost to undertake this type of screen depends on how it will be performed. If we assume that all customers are screened (but that any customer would only be screened once no matter how many times they order in a year) and we assume that entering a name into a database and waiting for an answer takes two minutes, then we can estimate that the 200,000 US CSNA customers would engender a total of 6,500 hours (400,000 minutes) to be screened in this manner, system-wide. This figure implies that just over three FTEs would be required to perform this screening, system-wide, which amounts to less than $1 million per year even at very high loaded rates for screeners. Clearly, since there are more than three CSNA companies, screening would likely require far more people who are not dedicated to this task if screening occurs at the CSNA companies themselves. Furthermore, if researchers who are probably legitimate are exempted from this screen (such as those from and with shipping addresses consistent with Merck's research headquarters, Dartmouth Medical School or USAMRIID), the cost would drop, as would be the case for systems that exempted long-standing customers from screening.

By far the most expensive customer screening method is licensing. To estimate costs of a CSNA licensing scheme based on the Select Agent Rule, we tried to find how much the government spends on implementing the Select Agent Rule. Unfortunately, it was difficult to determine how much the CDC spends on the Select Agent Rule because these costs are rolled into other biosafety/biosecurity efforts. We did determine that USDA-APHIS, spends roughly $3 million to implement the Select Agent Rule at its 73 facilities.[273] As discussed in Chapter 4, there are a total of 406 facilities cleared to work with select agents, so the total spent on implementing the Select Agent Rule is approximately $15 million/ year. Since we do not know how many facilities would need to be cleared for CSNA use, we will base the estimate on the licensing costs for a CSNA system on personnel that need clearance. Since there are 8,000 personnel cleared to work with select agents and roughly 200,000 scientists who work with CSNA, we can estimate that a CSNA inspection regime based on the select agent rule would cost roughly $375 million to implement. Alternatively, if the system licensed only customers who intend to order sequences related to pathogenicity (which we estimate at roughly 20,000 researchers in the US), then a similar system would cost only $40 million/year. Clearly, since biosafety inspections as rigorous as those required by the Select Agent Rule are unnecessary for almost all CSNA orders, this estimate is an upper limit. Also, the cost of licensing foreign customers of CSNA suppliers, by investing authority in foreign licensing boards or sending CDC inspectors overseas, for example, must be considered.

If the CSNA licensing system is based instead on radiation licenses, the costs are lower. The NRC itself requested $47 million to perform inspections, respond to incidents and undertake rulemaking activities.[274] Because the NRC performs inspections only in non-agreement states which are about a dozen states with little high-technology industry, it is difficult to estimate how much is spent nationwide on inspections. To approximate costs for inspections nationwide, we will use the state of Maryland as an example. We assume that Maryland performs an average quantity of radiation-safety inspections (it has a reasonably large high-technology and research base—states such as Massachusetts, New York and California would have more radiation-safety activities, states like Alabama and Idaho would have fewer). Maryland spends $2 million/year on their radiological health program[275], so we can estimate that $100 million/year is spent nationwide.

---

[273] Based on FY07 funding, as referenced in Franco, C and Deitch, S "Billions for Biodefense: Federal Agency Biodefense Funding" *Biosecurity and Bioterrorism* **5** (2), 2007.

[274] US NRC Performance Budget, Fiscal Year 2008, NUREG-1100, Volume 23.

[275] Based on the request for FY08, from the 2007 MDE Fiscal Analysis Project Draft Budget, available at: http://www.mde.state.md.us/assets/document/FAP_Meeting_07-16-2007.pdf

As discussed above, the cost of a CSNA licensing regime could be reduced if it builds upon other licensing regimes. For instance, if a license to obtain radionuclides is taken as a hallmark of being a legitimate researcher (because, in agreement states at least, a government official inspected the workspace), then all those already possessing these licenses could be authorized to receive CSNA without further clearance needed. Due to the substantial overlap between laboratories working with radioactive substances and CSNA, this measure would eliminate the majority of customers who need to be newly granted a license. It should be noted however, that the ability to handle radioactive material does not necessarily translate into the ability to handle pathogens that one may create from CSNA sequences, so the biosafety benefit of this option, as mentioned above, is dubious.

As mentioned in the description of the notional system, above, even if screening is performed entirely at the CSNA providers, if government-directed screening standards exist, the government must bear costs associated with evaluating and ensuring compliance to the screening regime.

Sequence Screening

The cost of screening sequences can be divided into three components, the cost to develop the screening tools, the cost to develop and curate a database of suspect sequences, and the cost of doing the screening itself. The cost of investigating false hits is estimated in the Confirmatory Screening section below. The cost of further development will be relatively minimal as tools already exist that are in use by CSNA companies. Some CSNA suppliers have, without government support, developed their own screening tools. Tools developed by a third party, like BlackWatch, could benefit from further evaluation and development, but the costs for this development should not exceed $1 million as the algorithms and interface are already proven. To develop the database of suspect sequences, a full risk and process assessment, which consider previous Material Threat Assessments, must be accomplished. Based on our experience performing similar studies, a thorough risk and process assessment to guide database development should cost about $500,000.

The cost of performing the screening itself varies depending on the screening interface. If the screening program can automatically pull sequences from electronic orders or from the instructions given to synthesizers, the actual cost of screening will be minimal. If a person must read orders and input the sequence by hand into a program, the costs increase dramatically. Here, we assume that enough computing power is available so that no backlog of requests to be processed develops, and that the average 25nt sequence could be entered in about two minutes. Given that about 30 million sequences are ordered yearly, a million work-hours (or 500 FTEs) would be required simply to perform primary screening of CSNA orders. If we assume a modest loaded rate for these screeners of $50,000/year (which would cover management and overhead costs), the cost simply to enter sequences into a screening tool and report hits would be about $25 million/year. If sequences must be entered by hand, the further problem of recruiting and retaining workers for this crushingly boring activity must be addressed.
The cost for sequence screening considers that all CSNA orders are screened. If only long oligos and genes are screened, we could eliminate about 75 percent of the screening volume (although the average length of a sequence, and the time required to enter the sequences by hand, would increase). Similarly, if only genes are screened, screeners must contend with vetting only 50,000 orders per year. It is inconceivable that a screener would have to enter 1,000 bases into a system by hand, and would more likely be able to cut-and-paste and electronic sequence into a screening program. Based on the same assumptions above, primary screening of only genes would cost approximately $50,000 a year.

Confirmatory Screening

The cost of confirmatory screening is determined by the false hit rate in primary screening, the method chosen in primary screening and the method chosen to investigate orders flagged in primary screening. Given the myriad choices available here, we can produce a rough estimate of confirmatory screening if we assume that each hit requires one hour to investigate, we assume that most legitimate orders for pathogenic sequences are not flagged due to customer screening efforts, and we assume a false hit rate similar to that expected from the use of BlackWatch results in one false hit in 10,000 orders.[276] Given 30 million CSNA orders a year, an error rate such as this would engender 3,000 hours of confirmatory screening, or the effort of just under two FTEs (or less than $100,000). Screening only long oligos or genes would drop the burden of secondary screening even further.

Funding the System

The costs of a CSNA oversight system could be borne by either the taxpayers or the CSNA customers, themselves. Note that customers could still bear the cost of the system even if the government performs the screening because CSNA orders could be loaded with user fees, such as those that support radiation safety inspections and licensing in Agreement States or attached to airline tickets to support screening at airports.

If customers must bear the cost of the screening system, how much are they wiling to pay before they will look elsewhere for suppliers? As mentioned in Chapter 3, most industrial customers of the CSNA industry are not primarily concerned with price, but academic customers would "start to look for other suppliers" if CSNA costs increased by about 25 percent. Therefore, a 10 percent cost increase would likely be a safe estimate for the additional costs that customers are willing to bear. Given that we estimate the worldwide market for CSNA products (which US suppliers dominate) is approximately $1 billion/year, customers could support a system that costs as much as $100 million/year. These funds seem sufficient to cover all possibilities except a licensing regime that would require inspections as thorough as those used for the Select Agent Rule.

If the taxpayer is to bear the costs of this system, we must compare the costs from similar biosecurity/biosafety regimes. Clearly, the most relevant comparisons are to the Select Agent Rule and the control of radionuclides. Assuming that these systems are viewed as cost effective by the taxpayer and that a CSNA screening system would not provide "more" security or safety than these systems, we can expect the taxpayer to bear similar costs. Therefore, we assume that the taxpayer should be willing to pay between $15 million/year, which we estimate as the cost for implementing the Select Agent Rule, given that the Select Agent Rule protects the taxpayers against a more proximate risk[277] and $47 million/year, which we estimate as the cost to the taxpayer for implementing radiation safety inspections,[278] which covers a similar user community.

[276] Since BlackWatch includes considerations for span and score, the actual false hit rate will likely be lower.

[277] That is the use of a select agent itself compared to the use of reagents to make a select agent that is then used.

[278] This considers only the NRC's budget for inspections and rule-making related to the control of radioactive substances, not user fees collected in Agreement States.

Conclusions

Several critical choices face those designing an oversight regime for the CSNA industry. The first choice is fundamental: which products of the CSNA industry should be regulated? In order to prevent hostile actors from synthesizing genes and viral genomes de novo, the system should regulate genes and oligos greater than 35nt in length. In order to prevent the magnification of genes from medical or environmental samples via PCR by hostile actors, the system should include oligos of any length. If only long oligos and genes are included in the system, most of the CSNA sold daily, most customers and some CSNA producers would not need to be included in the system. Also, because short oligos are more numerous (in fact accounting for 80 of the orders) and more likely by chance to match a suspect sequence, the exclusion of short oligos from the system is likely to dramatically reduce the false hit rate in primary screening. Furthermore, CSNA products can be DNA analogs or chemically modified DNA, or come attached to substrates. Most of these non-standard-DNA products are not ideally suited for use in PCR or for viral or gene synthesis and would only be used in the illicit production of a biological agent if the weaponeer feared identification by the oversight system. Excluding non-standard oligos would have a minimal benefit on reducing the total quantity of orders that must be screened, but would have the benefit of putting several small companies outside of the oversight regime. Final decisions on which CSNA types and formats should be beyond regulatory reach should be based upon a full weapons-process assessment.

The second main question to resolve is: what group will provide oversight and perform the screening? Screening performed by the CSNA providers themselves would not only be more palatable to the industry itself due to the retention of proprietary information about CSNA orders but also because the system would likely screen efficiently enough, due to the fact that the screeners and suppliers are co-located, to enable the supply of oligos in a time-frame expected by their customers. In contrast, screening performed by a single, national entity would ensure compliance, be less likely to be circumvented by an adversary and facilitates the use of data from ongoing screening to improve future screening. However, this system may be too cumbersome to meet the time demands of the oligo industry and would require the transmittal of proprietary business information.

There are several options for screening an order, including screening the customers only or screening customers in the context of the sequences they order. Because we estimate that approximately 1,000 orders are made daily by legitimate researchers for CSNA sequences associated with pathogens, customer information must be included in any screening system. Customers could be screened by information in their e-mail address, by screening against lists of known terrorists or denied parties or by ensuring that the customer possesses a license to obtain CSNA. These measures would be effective at preventing hostile actors from acquiring critical reagents but would have little biosafety benefit. In contrast, if licensing is combined with sequence screening, some customers could be issued licenses to obtain CSNA from pathogens; those without this type of license could only obtain CSNA from non-pathogenic organisms, thereby improving biosafety. A licensing system could be implemented in a variety of ways, including grandfathering in those who have ordered CSNA products in the past. However, it is unclear how licensing would be implemented for foreign customers of US companies. Sequence screening could be performed by software that is currently available, such as BlackWatch, this software could be improved or entirely new software could be developed. Details on what CSNA orders should be screened against must rely on a detailed weapons process assessment. Any hit should be confirmed before an order is denied or a field investigation is initiated. This confirmatory screening could involve a phone call to the customer's biosafety officer, or determining if the sequences ordered match their previous orders better than pathogens of concern.

Although it is clear from existing practices that questions of biosafety should be investigated by the CDC or USDA, questions of biosecurity could be investigated by the FBI or DHS's HITRAC. Additionally, standard practices need to be immediately implemented even if no single screening system is adopted. CSNA producers must be supplied with a point of contact at the USDA-Animal and Plant Health Inspection Service, CDC, the FBI (the WMD coordinator of the local field office) and/or HITRAC. Furthermore, standard procedures must be given to the CSNA supplier as to whether or not to fill the orders referred to a field investigation, and if the order is not to be filled, what to tell the customer.

Furthermore, to support forensic investigations, information on orders should be stored. If these orders are stored at the CSNA producers and only inspected when needed, forensic investigations would be slowed but proprietary information would be protected. If order information (or information only on denied orders or orders that failed primary screening) is stored in one location, this database would be easier to search in times of need but may raise concerns dealing with proprietary business information.

Finally, a CSNA oversight system could be underpinned by a variety of mechanisms. Self-regulation is, by its nature, the most palatable to industry, but is likely to be inconsistent and leave gaps to be exploited by adversaries. Voluntary federal standards could be adopted to improve adherence to the system industry-wide, and may be palatable to industry because of their desire for guidance from the federal government and the possible reduction in vulnerability to tort claims. Guidelines that dictate that any institution receiving NIH (or other government) funding can order CSNA only from suppliers that employ a standard screening system would improve compliance to the system by industry in the US and abroad by controlling the buying patterns of a large part of the customer base. Furthermore, companies could not escape regulation by relocating production facilities overseas. Lastly, screening could be compelled by formal changes in US law or regulations, such as the Select Agent Rule and the Commerce Department Regulations.

It is important to consider that the most effective oversight system may actually be multiple systems in one: one tuned to provide oversight to the gene industry and one to the oligo industry. We realize that the requirements of the customers of the gene synthesis, compared to the oligo synthesis, industry are vastly different as are the time and financial constraints faced by both industries. Furthermore, an oversight system may be implemented by including components of the CSNA industry one at a time. Because many gene synthesizers already perform screening, because the quantity of genes made pales in comparison to the quantity of oligos made, because genes are probably more enabling to adversaries than oligos, and because the cost and time constraints of gene production are more relaxed than oligo production, implementing regulation by starting with the gene synthesis industry seems appropriate.

The cost of a CSNA screening system could almost certainly be covered by fees attached to CSNA products. From talking to customers, we estimate that the market would be able to raise approximately $100 million a year in user fees without disrupting the buying patterns of customers. This amount of funding would support most types of screening systems except ones that exactly mirror the rigors of the Select Agent Rule.

Appendix I. CSNA Industry Participation in Interviews

Gene Companies (n=11)		
Companies we spoke with	***Companies that declined to speak with us***	***Companies we were unable to reach***
Blue Heron Biotechnology	Celtek Genes	
Codon Devices	Genecopoeia	
Coda Genomics		
Commonwealth Biotechnologies		
DNA 2.0 Inc.		
Eton Bioscience		
GenScript Corp		
MCLAB		
NextGen Sciences		
Oligo Companies (n=20)		
Companies we spoke with	***Companies that declined to speak with us***	***Companies we were unable to reach***
Azco	Bio Search Technologies Inc.	Ambion
BioServe Biotechnologies LTD	Dharmacon	LC Sciences
Combimatrix	Molecula	Primm Biotech
Gene Tools LLC	Omega Bio-Tek	Tri Link Biotechnologies
Fidelity Systems	PrimeSyn Lab Inc.	
Invitrogen		
The Midland Certified Reagent Company, Inc		
Ocimum Biosolutions		
Oligo Factory, Inc.		
Operon Biotechnologies		
Sigma-Genosys		
Gene and Oligo Companies (n=9)		
Companies we spoke with	***Companies that declined to speak with us***	***Companies we were unable to reach***
ChemGenes Corporation	Epoch Bio Labs	Bio-Synthesis Inc.
Gene Link		Bionexus
Integrated DNA Technology		MWG Biotech Inc.
Syn Gen Inc. (fastaDNA)		Retrogen Inc

Appendix II. Redacted due to commercial proprietary information