S&T Developments of Relevance to the Biological Weapons Convention

United States of America

I. Introduction

Since the previous review of scientific and technological developments in 2011\(^1\), no new discoveries or technologies have fundamentally altered the nature of life sciences research or raised questions concerning the scope of the Convention. However, the pace of developments identified in 2011 has continued to accelerate, expanding the possibilities not only for benefits to public health and medicine, agriculture, the environment, and other peaceful applications, but also for misuse for purposes banned by the Convention. Individual scientific discoveries and technological tools continue to grow more accessible, to develop more rapidly, and to converge with other disciplines such that the threat from biological weapons and our ability to respond effectively is evolving rapidly. This paper highlights several major advances and trends in the life sciences over the last decade and examines their implications for the Convention.

II. Specific S&T developments since 2011

Advances in genome and gene editing technology

Altering the genetic material of natural organisms is not new. Humans have been doing this for thousands of years through selective breeding of crops and livestock. However, 20\(^{th}\) century insights into heredity, followed by the ability to directly manipulate genetic material have enabled the development of “biotechnology” – technology that applies to and/or is enabled by life sciences innovation or product development, for example, the intentional harnessing of genes and other cellular or biomolecular processes to create technologies, medicines, and products.

Since 2011, scientific and technology developments continued to advance the accuracy, speed, and understanding of genome editing for a wide range of beneficial applications. CRISPR/Cas technology is a notable development that led to a Nobel Prize in 2020 and is used around the world\(^2\). This genome editing tool and others like it enable research in the life sciences to be carried out with a level of precision, efficiency, and scope that was not previously possible. Further, the advances in precise genome editing technologies are creating new or improving existing clinical therapies. For example, clinical data is demonstrating how genome or gene editing tools may be able to cure patients with diseases with known genetic disorders (like sickle cell disease) or to target specific parts of a patient’s body (like cancer cells)\(^3\). Beyond the health and medical fields, genome editing technologies can be and have been used to precisely

---

\(^1\) BWC/CONF.VII/INF.3 - New scientific and technological developments relevant to the Convention – Background information document submitted by the Implementation Support Unit


introduce desirable traits for sustainable and resilient food production – like new oilseed varieties that have higher levels of omega-3 fatty acids or grain crops that are more tolerant to drought. Similarly, genome editing technology makes it more feasible to convert micro-organisms into cellular factories for desired products (i.e., metabolic pathway engineering).

**Advances in vaccine technology**

Vaccines protect against the spread of disease by providing immunity to individuals. Since the vaccines for smallpox, cholera, anthrax, and plague were developed over a century ago, vaccine technology has continued to improve and be a cornerstone for disease prevention and control. Historically, vaccines have taken approximately 10-20 years to develop, but in 2020 multiple safe, effective, and high-quality vaccines against a novel virus (SARS-CoV-2) were researched, developed, and manufactured at scale in just one year from when the outbreak was reported, by leveraging decade-long basic research in a number of fields. A combination of complementary strides in the scientific knowledgebase, DNA synthesis technology, vaccine research using structural biology tools, advanced manufacturing, bioinformatics, longstanding investments in biological defense research and development, and international collaboration all enabled this truly remarkable achievement.

One specific scientific development that went beyond accelerating the traditional vaccine technologies, such as inactivated or attenuated viruses, was the advent of mRNA- and viral vector-based vaccines. These vaccine platforms bypass the need to pre-produce and package materials like proteins by directly delivering molecular *instructions* into patients whose own cells in turn produce the needed antigens to build immunity. This process also enables the use of manufacturing technologies that also appear promising for rapidly addressing other emerging biological threats. Prior to COVID-19, no vaccines had been authorized using mRNA vaccine technology and only a handful of viral vector vaccines (for Ebola in humans and for several veterinary diseases) were approved. One key safety and security benefit of this vaccine technology is that the manufacturing process removes the risks associated with handling, and growing large amounts of, live or attenuated virus, as vaccine development can be performed without those materials. These technologies are now being explored for a variety of other diseases, such as HIV, influenza, and Nipah virus and it is theoretically possible that the technology could be applied to vaccines for childhood immunizations, other microbial pathogen threats, and to treat noncommunicable diseases including cancer. Furthermore, such “platform technologies” may be readily adaptable for use if a novel pathogen emerges as a public health threat.

**Advances in genomic sequencing**

Genome sequencing and especially whole genome sequencing – the ability to read the entire genetic instructions of an organism or virus – is increasingly important for a broad spectrum of applications. For instance, sequencing is used during the public health response to a disease outbreak to detect and track mutations, for development of medical treatment tailored for a specific person suffering from cancer or other conditions with a genetic basis (“precision

4 https://www.immune.org.nz/vaccines/vaccine-development/brief-history-vaccination
medicine”), and for environmental conservation efforts by identifying and monitoring invasive species.

**Disease outbreak response**

Genomic sequence data from samples taken from patients or the environment can be used to detect outbreaks and novel zoonotic or other pathogens that pose a risk to health security, support efforts to develop vaccines and diagnostics, support decision making for non-pharmaceutical interventions, and carry out ongoing surveillance to identify new variants. The use of genomic sequencing to advance public health has grown rapidly, particularly during the COVID-19 pandemic.

Since genome sequencing works by reading the genetic material in a sample, sequencing can be widely used to identify novel pathogens that were not detected by agent-based diagnostic methods. For example, genome sequencing was used to identify SARS-CoV-2 – a new virus – as the causative agent of a cluster of unknown pneumonia cases in Wuhan, China in early 2020. Knowing the genome sequence of the virus spurred the rapid development of diagnostics, therapeutics, and vaccines that are now being used in countries worldwide.

Ongoing genomic sequencing of patient samples linked to data regarding surges in cases or clinical outcomes during a disease outbreak can track the emergence of new variants, aid in risk assessments, and monitor how they spread. During the COVID-19 pandemic, sequencing paired with other data enabled the identification of new variants of concern globally, allowing countries to identify, assess, and track the spread of variants with different properties, such as the initial Alpha variant and successive variants such as Omicron. This approach has allowed countries to implement changes in public health measures and modify existing diagnostic tests and medical countermeasures, as needed.

**Precision medicine**

The wide availability of genome sequencing allows doctors and researchers to more accurately predict which treatment and prevention strategies for a particular disease will work for specific groups or individuals based on their genetic variabilities. This approach contrasts with the traditional “one-size-fits-all” approach, in which disease treatment and prevention strategies are developed for the average person, with less consideration for the differences between individuals.

Cancer is the target disease of some of the most promising precision medicine approaches available today. Cancer usually comes about through the gradual accumulation of genetic changes (mutations) in genes that control cell growth. Depending on where in the body the cancer arises and the types of genetic changes the cells accumulate, different types of cancer can have very different genetic profiles and respond differently to treatments. By comparing the DNA from a patient's tumor to that of their normal cells, researchers can learn how the cancer
came about and what treatments might be most effective\textsuperscript{5}. By tracking the genetic profiles of their patients' tumors, doctors can learn which treatments work best for which patients.

**Advances in information processing**

The life sciences have greatly benefited from new computing technology that enables rapid processing of large quantities of data with techniques ("machine learning") that allow progressively more accurate conclusions to be drawn. Related advances in computational modeling have been used for predicting the spread of disease outbreaks and the impact of restrictions on disease spread, as well as associated impacts on transportation, local economies, and other factors. Further computational modeling provides a means of rapidly identifying promising pharmaceutical candidates.\textsuperscript{6} Two recent examples relevant to the Biological and Toxin Weapons Convention are predictions of protein structure and attribution of genetic engineering.

**Protein folding**

Proteins are chains of amino acids that are present in all living organisms and include many essential biological compounds such as enzymes, hormones, and antibodies. Proteins form a three-dimensional shape that is necessary to their function. Predicting the shape of a protein is critical to understanding exactly how it carries out its task and how this action may be modified, blocked, or enhanced in order to treat disease.

Unfortunately, the number of different shapes that a given protein could take based on its amino acid sequence is astronomical. Until recently, scientists have depended on expensive and time-consuming laboratory methods to determine the structure of a protein experimentally. Such efforts have identified the structures of about 170,000 proteins over the last sixty years, out of the approximately 200 million proteins estimated to exist in nature\textsuperscript{7}.

In August 2022, the artificial intelligence company DeepMind unveiled the likely structures of 200 million known proteins, from organisms ranging from bacteria to humans. This remarkable result comes from the artificial intelligence program AlphaFold, which has substantially advanced the protein-folding field by more accurately predicting the three-dimensional shapes of proteins from their amino acid sequences than previous technologies\textsuperscript{8}. Many applications will still require additional experimental evidence to validate these structures, including to aid in rapid pandemic response and other drug discovery efforts. However, this development provides a significant head start for discoveries. The predicted protein structures were released into an existing free database through a partnership with the European Molecular Biology Laboratory’s European Bioinformatics Institute.


**Genetic engineering attribution**

The rapid development of techniques for synthesizing genetic material or editing naturally occurring sequences has raised concerns that such genetic engineering techniques might be misused to produce a biological weapon, for example, by making a naturally occurring pathogen strain more infectious or more survivable in the environment. Addressing these concerns poses two separate challenges – detecting that a genetic sequence has been engineered and determining who was responsible ("attribution").

The FELIX project in the United States is developing new experimental and computational tools to identify indicators of genetic engineering. Under the UN Secretary-General’s mechanism for investigating alleged biological attacks, efforts are underway to develop an international laboratory network that would be capable of detecting modified genetic sequences in pathogens using specialized information processing tools.

Considerable progress has been made in developing tools for attribution of genetic engineering. In 2020, more than 300 teams from around the world took part in a data science competition to identify the laboratory of origin of engineered genetic sequences with the highest possible accuracy. Pulling from open-source published sequences, top accuracy scores exceeded the previous state of the art. Winning teams adopted a variety of different technical approaches, demonstrating the diversity of information processing methods that can be applied to genetic engineering attribution, as well as the potential for new machine learning approaches to further improve on existing tools. All computer software from the winning submissions will be made publicly available.

**Advances in the synthesis of biological molecules**

Prior to the 1970s, biological materials needed to be obtained from living animals, plants, or other organisms. The recombinant DNA revolution made it possible to directly synthesize such materials without needing to collect, grow, or raise the original organism. One example is antibody production for pharmaceutical purposes, such as antibody treatments for rheumatoid arthritis (like Adalimumab). While antibodies can be harvested from infected animals, many can now also be manufactured directly by human or other cell lines in the laboratory which can lower the products’ cost, increase manufacturing speed, and minimize patient allergic reactions. Over the past decade, the synthesis of biological molecules has become less expensive and more widespread, and the type of products and the ways to produce them has continued to expand. Biological molecules like DNA and other nucleic acids and proteins can be chemically synthesized in longer lengths and lower costs around the world than ever before. Many companies now offer custom nucleic acid synthesis.

Advances in nucleic acid synthesis technologies have facilitated easy-to-use, efficient, and accurate systems to study viruses and how they cause disease, to understand what new mutations may mean for virulence or the ability to evade immune responses, and to standardize the

---

9 https://www.iarpa.gov/research-programs/felix
10 https://altlabs.tech/geac/
genomic blueprint from which to grow virus stocks for experimentation, therapeutic, and vaccine development. At the beginning of the COVID-19 pandemic, nucleic acid synthesis technology enabled the SARS-CoV-2 virus to be synthesized within a month after the genome sequence was published on January 11, 2020.

Other biological molecules like lipids, large proteins, or biochemicals can be synthesized directly or in specifically engineered organisms to manufacture certain products. These products can range from medical, agricultural, and other commercial applications, such as the manufacturing plant-based heme (soy leghemoglobin) in yeast to flavor plant-based foods11 or bioplastics (polyhydroxyalkanoates) that are potentially more eco-friendly alternatives to traditional plastics.

III. Trends in Science and Technology

In addition to specific scientific and technical advances, such as those highlighted in the previous sections of this paper, there are more general trends in science and technology that individually and collectively are relevant to the Convention. Among these are: systems biology, which combines information from several specialized approaches; greatly enhanced international collaboration; rapid accretion and assimilation of data; creation of platforms for open sharing of data; and rapid open publication of results.

By combining information gathered from different highly specific approaches (“omics” methods), researchers are able to obtain a much better understanding of how biological systems function, interact, and are in fact interdependent. This can provide insights into basic biology, mechanisms of disease, and potential drug targets. For example, researchers are applying these approaches to help identify treatment candidates that work across multiple coronaviruses12 or to help identify the genetic basis for heart disease13. The desire to better understand the complexity of biological systems, accelerated by the urgency of responding to the COVID-19 pandemic, is fostering greater interdisciplinary research and collaboration within and among countries. Key examples are the CEIRR Network14 (Centers of Excellence for Influenza Research and Response Network), a research network created to study influenza and combat influenza outbreaks, and the CREID Network15 (Centers for Research in Emerging Infectious Diseases), a coordinated group of research centers situated around the globe where emerging and re-emerging infectious disease outbreaks are likely to occur.

The rapid pace of research and creation of large sets of data is leading to repositories that enable this information to be more accessible. During the COVID-19 pandemic, platforms like GISAID16 allowed public accessible, real-time sharing of sequences and NextStrain17,

11 https://faq.impossiblefoods.com/he/en-us/articles/360034767354
14 https://www.ceirr-network.org/
15 https://creid-network.org/
16 https://gisaid.org/
17 https://nextstrain.org/
GenBank\textsuperscript{18} and other international nucleotide sequence database collaborations facilitated open, real-time sharing of genomic information for SARS-CoV-2 sequences, allowing researchers and public health officials to better track the course of the pandemic and identify variants. In addition, the Protein Data Bank (www.wwpdb.org) was populated with hundreds of SARS-CoV-2 protein structures from x-ray centers, which allowed targets to be identified for therapeutics design. Data repositories also help computational biologists develop tools for processing large amounts of data in order to extract new insights, such as changes in the composition or characteristics of circulating SARS-CoV-2 viruses, and impact of any anticipated changes in public health response strategy at local, national, or global level.

Traditionally, research results were published in technical journals after careful review by panels of experts (“peer review”) and access to the journal articles required a subscription. The need for a rapid response to the COVID-19 pandemic, however, greatly increased the need and pressure for greater and immediate public accessibility to research results. In response, many journals opened access for COVID-19-related publications. Sites for sharing draft research articles were developed, which although not peer-reviewed, enabled more widespread use of data in-time sensitive pandemic response and public health decision making. Although this development allowed knowledge to be shared more rapidly and widely, the lack of quality control meant that researchers and public health officials needed to be especially alert for poor or inaccurate work.

Although the COVID-19 pandemic fostered greater international collaboration, sharing of data, and access to research results, it also stimulated interest in strengthening national capabilities for research and development related to emerging and high-consequence pathogens. For example, numerous countries announced plans to build new high and maximum containment laboratories. Such laboratories require extensive safety and security measures. Not only are they expensive to build, but they also require dedicated and sustained funding and expert staffing to ensure that they are safely and properly operated, maintained, and secured.

**IV. Implications for the Convention**

Individually, each of the advances described in this paper makes a substantial and growing contribution to our efforts to understand how living organisms work. Applying this knowledge for peaceful purposes is yielding substantial and growing benefits for humankind, including improving public, animal, and environmental health, as well as producing better and less expensive food. Collectively, the impact can only be described as revolutionary. It cannot be overemphasized that improving the ability to prevent, detect, and respond to natural outbreaks of infectious disease also helps to build defenses against both accidents and the deliberate use of disease as a weapon. A strong public health system prepared to prevent, detect, and respond to infectious disease threats can help to deter the use of biological weapons by reducing their potential effectiveness and impacts, including in the event that a biological weapon is used, to minimize the deaths, injuries, and economic damage that may result.

On the other hand, it has long been recognized that advances in understanding living systems can be used not only for beneficial purposes, but also to develop ways to disrupt those systems; such knowledge is “dual use.” For this reason, the Convention prohibits those actions that have no justification for peaceful purposes.

\textsuperscript{18} GenBank Overview (nih.gov)
The advances described in this paper are each “dual use” in character. Any of them could be misused for biological weapons purposes. The comprehensive scope of the Convention covers these advances, but continuous review is necessary. In particular, in planning and conducting life sciences research, careful oversight is needed to preserve the benefits of such research while minimizing the risk of misuse of the knowledge, information, products or technologies provided by such research.

In summary, the actual and potential benefits of the advances in science and technology described in this paper are numerous, but risks do exist that need to be carefully monitored, including through creation of a BWC science and technology advisory mechanism, so that, if necessary, appropriate safeguards are in place.