EDITORIAL

Check for updates

Protein design meets biosecurity

The power and accuracy of computational protein design have been increasing rapidly with the incorporation of artificial intelligence (AI) approaches. This promises to transform biotechnology, enabling advances across sustainability and medicine. DNA synthesis plays a critical role in materializing designed proteins. However, as with all major revolutionary changes, this technology is vulnerable to misuse and the production of dangerous biological agents. To enable the full benefits of this revolution while mitigating risks that may emerge, all synthetic gene sequence and synthesis data should be collected and stored in repositories that are only queried in emergencies to ensure that protein design proceeds in a safe, secure, and trustworthy manner.

Nature's proteins elegantly address the challenges faced during the slow march of evolution, but today's problems, such as global pathogens, neurodegenerative diseases, and ecosystem degradation, require new solutions. AI-accelerated protein design can help tackle many of these issues. Machine learning–based methods enable the fast creation of biomolecules with diverse structures and functions that often have no detectable sequence homology to any known proteins. Concurrently, exponential improvements in DNA synthesis cost, quality, and speed have simplified encoding these proteins into synthetic genes. Last year, the first drug developed through computational protein design, the COVID-19 vaccine SKYCovione, was approved internationally. Many more such innovations are possible with this approach—and on short order. But as reflected in last year's global AI Safety Summit in the United Kingdom, the road to regulating AI is likely to be long and complicated. Progress in computational protein design could be hindered by overly restrictive AI regulations. The good news is that AI tools for protein design are highly specialized, and hence risk mitigation should be more straightforward.

Prior to the 2023 AI Safety Summit, a conference in Seattle, Washington, convened international representatives from academia, industry, philanthropy, and government agencies to discuss AI-enabled protein design, particularly for pandemic preparedness and drug development. The manufacture of synthetic DNA was recognized as a key biosecurity control point. Among the recommendations that emerged from the meeting was a policy of screening and logging all synthesized genetic sequences. This would present a practical barrier to the creation of harmful biomolecules, whether accidental or intentional.

> "Enhanced security need not threaten information sharing..."

Since 2004, the regulation of DNA synthesis, proposed and then voluntarily adopted by members of the International Gene Synthesis Consortium (IGSC), has been widely practiced in academia and the biotechnology and pharmaceutical industries. Currently, requests to academic, private, and government institutions for DNA sequences are screened by the IGSC for homology to pathogen components on a consensus list.

Going forward, these checks could be linked with the synthesis itself—whether chemical or enzymatic—such that each synthesis machine requires cryptographic short exact-match searches for each new genetic sequence. Screening sequences alone may not be sufficient because proteins generated through de novo design may have little or no sequence similarity to any natural proteins, complicating homology detection. Hence, there is a need to log synthesized sequences, using encryption as necessary to protect trade secrets. If a new biological threat emerges anywhere in the world, the associated DNA sequences could be traced to their origins. A "selective revelation" policy could ensure that such queries occur only under exceptional circumstances and on the basis of preestablished criteria. As biological complexity makes it highly unlikely that a dangerous agent could be created in one attempt, this capacity to trace nascent threats to their origins should be effective. Besides providing an audit trail, awareness that all synthesized sequences are being recorded may deter bad actors. Screening and logging practices should be standardized, practiced internationally, and extended to benchtop nucleic acid synthesizers.

This protein design security strategy depends on input from all relevant communities to support the required infrastructure and define the human, institutional, and governance requirements. Ideally, an international group such as IGSC should take the lead but work with governmental and nongovernmental organizations. Enhanced security need not threaten information sharing or transparent communication, the hallmarks of modern science; the use of biosecurity as an excuse to not share new methods and advances should be discouraged by science funders, publishers, and policy-makers. Rather, security in this fast-moving field should be framed as maximizing progress to address pressing societal concerns.

–David Baker and George Church

David Baker is a professor in the Department of Biochemistry and director of the Institute for Protein Design at the University of Washington, Seattle, WA, USA. dabaker@uw.edu

George Church is a professor of genetics at Harvard Medical School, Boston, MA, USA. gchurch@genetics.med.harvard.edu

10.1126/science.ado1671