ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus
Biologies

Patrick Berche

The dangerous biology of pathogenic germs

Volume 347 (2024), p. 77-86

Online since: 19 September 2024

https://doi.org/10.5802/crbiol.157

The Comptes Rendus. Biologies *are a member of the Mersenne Center for open scientific publishing*
www.centre-mersenne.org — e-ISSN : 1768-3238

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Biologies
2024, Vol. 347, p. 77-86
https://doi.org/10.5802/crbiol.157

Review article

The dangerous biology of pathogenic germs

Patrick Berche[a]

[a] Emeritus Professor at the Université Paris Cité, Member of the Académie Nationale de Médecine, France

E-mail: patrick.berche@gmail.com

In memoriam of Professeur Henri Korn

Abstract. The convergence of biotechnologies with other disciplines, including computer science and Artificial Intelligence (AI), may make it possible to carry out dangerous genetic manipulations on pathogenic germs, as the gain-of-function experiments exacerbating virulence, as those carried out on myxoviruses and coronaviruses. Moreover, it is now possible to chemically synthesise any microorganism from *in silico* sequences, including the most dangerous viruses (poxviruses, Ebola, etc.), whose sequences are accessible. It might even be possible to use AI to design new germs that could be used as biological weapons.

Keywords. Gain-of-function, Synthetic biology, Biological weapons, Biosecurity.

Manuscript received 15 May 2024, accepted 18 June 2024.

We are witnessing an unprecedented revolution in the biological sciences, marked by the convergence of biology with many other disciplines, as chemistry, physics, biotechnology, nanotechnology, mathematics, computer science and artificial intelligence, etc. The major technological impacts have been the rapid sequencing of DNA (1977), enabling the creation of data banks, the discovery of reverse transcriptase (1970), to copy RNA into complementary DNA (cDNA), and the chemical synthesis of DNA from *in silico* sequences.

It is now possible to modify or even create biological entities (genes, proteins, microorganisms, viruses, etc.) with new properties, mainly used for medical applications. Experiments inactivating certain genes of pathogenic germs through mutations or deletions allow the identification of those involved in virulence and contagiousness. Thus, it is possible to link the phenotype to the genotype of pathogens, which is useful for preparing vaccines and anti-infectious drugs. Conversely, virulence can be exacerbated by modifying or introducing certain genes involved in pathogenicity into other microorganisms, which is facilitated by CRISP-Cas9 technology [1]. These experiments are known as gain-of-function (GoF) or Dual-use research of concern (DURC). Although the majority of these experiments are designed for useful purposes, some pose problems of dual research and biosafety, with possible accidental escapes from the laboratory, especially when dealing with potentially pandemic pathogens (PPP) viruses, such as influenza viruses, coronaviruses or poxviruses.

Thanks to DNA sequencing, hundreds of thousands of complete pathogen sequences are available in databases. For example, 11,694 complete viral genome sequences have been deposited in the Viral Genome Resource [2], including all known pathogenic viruses.

1. Gain-of-function experiments

Many pathogenic germs that have been colonising wild animal species for a long time circulate silently in hosts that form viral reservoirs. Highly adapted to each species, they recognise specific receptors that enable them to colonise and penetrate host cells. When they encounter new hosts, they are generally

not infectious, as there is a species barrier preventing any colonisation. Crossing this barrier is rarely possible directly, as it depends on the frequency of interactions between animal species and their genetic proximity, but also on the plasticity of the microorganisms' genomes.

Pandemic germs are often RNA viruses with a high mutation rate. Their main wild reservoirs are birds, bats and rodents. In these animals, these viruses exist as quasispecies, which are swarms of several genotypes. The emerging wild virus is the most efficiently proliferating genotype [3, 4]. The transition to humans is a sequential process, usually requiring progressive adaptations via intermediate animal hosts in which adaptive mutations occur and accumulate [5]. These intermediate hosts have been identified for avian viruses (pigs), SARS-CoV1 from 2002 (civets, raccoon dogs, badgers), MERS-CoV (dromedaries) and Nipah virus (horses) [6]. Initial human contact with wild viruses is usually unsuccessful, resulting at best in sporadic cases of non-contagious infection or asymptomatic seroconversion [7]. This was the case, for example, with H5N1 avian influenza in humans [8] or the SARS outbreak in 2002, when almost 80% of the animals in Guangzhou markets showed anti-SARS-CoV1 antibodies [9].

Gain-of-function experiments can mimic this adaptation process of wild viruses. There are several ways of increasing the virulence and contagiousness of an infectious agent. It has long been known how to create random mutations that progressively increase virulence through passages in animals or cell cultures. One can also carry out genetic manipulation to induce mutations, through directed mutagenesis or insertion of genetic material. More recently, it was possible to create chimeric mutants by recombining certain pathogenic viruses, for example to modify their cellular tropism. Occasionally, an experiment can lead to an unexpected gain-of -function. In 2001, Australian researchers looking for ways to develop immunocontraceptive viral vaccines generated an extremely virulent mouse poxvirus (ectromelia). By adding the murine interleukin-4 (IL-4) gene, the recombinant virus caused mortality in all lines of mice, even after vaccination [10]. The new virus completely suppressed NK and cytotoxic T responses. Bearing in mind that the human smallpox virus is related to this virus and has the same genetic organisation, there is every reason to fear that a recipe for a formidable biological weapon was published. Was it necessary to publish these results?

Faced with recurrent threats of pandemics, due to avian flu (H5N1, H7N9), SARS (2002) and MERS (2012), researchers have tried to understand the molecular mechanisms of the pathogenicity of these viruses by studying the main virulence factors, such as the haemagglutinin HA of myxoviruses and the Spike protein of coronaviruses. The aim was to predict the emergence of a new virus, potentially better adapted to humans, from mutations created in certain putative virulence genes. Dangerous gain-of-function experiments were thus carried out on influenza viruses and coronaviruses.

1.1. *Influenza viruses*

The influenza virus, *Myxovirus influenza*, consists of single-stranded RNA (12–15 kb) harboring eight genes carried by separate segments, including the two virulence genes encoding haemagglutinin HA and neuraminidase NA, expressed on the surface of the virus envelope. When pigs are co-infected with several influenza viruses, reassortments can occur, leading to the emergence of new pandemic viruses. There are currently fears that a new influenza pandemic could emerge from the H5N1 avian virus, which triggered epizootics with human cases from 1996 onwards [11]. This virus recognises avian sialic acids (SA) (SA α-2.3 galactose), but does not recognise human SA (SA α-2.6 galactose) present in the upper respiratory tract, which explains why it is not contagious to humans. However, after deep inhalation in contact with poultry, the virus can reach the human bronchioles and alveoli where the avian receptor is present. This explains the high mortality rate and low contagiousness of the virus to humans. A pandemic alert was issued in Hong Kong in 2005, following clustered human cases without any real epidemic [12]. Between 2005 and 2024, there have been 888 cases of avian influenza in humans, including 463 deaths (52% mortality) [13]. However, it is feared that a few mutations could be enough to transform the H5N1 virus into a fearsome pandemic virus. In 2012, in an attempt to predict the risk of a new pandemic emerging, two teams sought to identify the crucial mutations in the HA gene that would enable the H5N1 virus to become contagious to humans by recognising human SA receptors. Researchers

in Madison replaced the H1 haemagglutinin gene of an H1N1 strain with the H5 gene of an avian H5N1 strain [14]. Unlike the H1N1 strain, the new virus is not contagious in ferrets. Random mutations in the H5 gene were created *in vitro* and then selected by adsorption onto turkey red blood cells expressing human SA. Thus, 370 mutants were obtained and individually screened to select those capable of binding human SA very efficiently. Of the nine mutants thus isolated, all had mutations in the receptor binding domain (RBD), in the 120–259 region of haemagglutinin H5. Four of which were contagious for the ferret inoculated by nasal route (N186K, S227N, Q226L, G228S mutations). These mutants cause mortality in 3–7 days after intra-tracheal and intra-nasal inoculations (10^6 viruses). The transmissible H5 reassortant virus preferentially recognized human-type receptors, replicated efficiently in ferrets, caused lung lesions and weight loss, but was not highly pathogenic and did not cause mortality. Another team, in Rotterdam (Netherlands), used directed mutagenesis to introduce two mutations in the haemagglutinin gene of an H5N1 strain (Q222L and G224S, to allow it to bind to human SA), and a mutation in the RNA polymerase gene (PB2) to make it active at 33 °C (the temperature of the human respiratory tract) instead of 40 °C (E627K). This H5N1 mutant, which is not airborne transmissible, was then nasally inoculated into ferrets to force adaptation through repeated passages (10 in total). Airborne-transmissible viruses were identified, all carrying two additonal mutations in the hemaglutinine gene (H103Y and T156A). This suggests that only 5 mutations (four in HA and one in PB2) may be sufficient to confer airborne transmission between mammals to the H5N1 virus. This H5N1 mutant was inoculated into ferrets via 10 nasal passages, resulting in a fatal and highly contagious airborne disease. Researchers identified four critical mutations in H5 (N182K, Q222L, G224S, N154K), enabling adaptation to mammals [15].

Noting the frequency of co-infections of pigs with avian H5N1 and human H1N1 viruses, a Chinese team reproduced a co-infection in cell culture to generate emerging pandemic strains [16]. They generated all possible reassortants between two strains: a H5N1 strain highly pathogenic for mice (carrying two mutations facilitating replication and SA binding in mammals) but unable to transmit efficiently through respiratory droplets in guinea pigs, and a H1N1 strain which is airborne transmissible in guinea pig but not highly virulent in mice. Out of the 127 hybrid virus, 35 were more pathogenic that the H5N1 strain. Among those, some were also capable of very efficient respiratory droplet transmission between guinea pigs. These dangerous experiments, published in 2012 and 2013 demonstrated that a few mutations enabling better replication and better binding to SA receptor were sufficient to cross the species barrier. These gain-of-function experiments led to a moratorium being imposed in the United States from October 2014 to December 2017. In October 2022, a major H5N1 epizootic occurred in mink farms in Spain, affecting almost 52,000 animals, with a mortality rising from 0.7% at the start to 4.3% after three weeks. Analysis of the viral genomes detected mutations in the replicative apparatus and the HA encoding genes that differed from those previously predicted, as well as additional mutations in other genes [17].

1.2. *Coronaviruses*

Coronaviruses consist of a single-stranded RNA of 26–32 kb with an envelope bristling with spicules made up of the S (Spike) protein. The genome of the 2002 SARS (Severe Acute Respiratory Syndrome) virus contains eleven genes. There are seven pathogenic coronaviruses in humans. Four (HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1) cause mild epidemic respiratory infections in children (rhinitis, etc.). Three others are responsible for severe respiratory infections: SARS-CoV1 in 2002–2003 (8346 cases, 646 deaths, mortality 7.8%); MERS-CoV from MERS (Middle East Respiratory Syndrome), which appeared in 2012 and remains rampant in the Middle East (>2000 cases from 2012–2024, 35%mortality); SARS-CoV2, responsible for the Covid-19 pandemic, emerging in December 2019 and evolving in iterative waves to date [18] (<14 million deaths, 0.6% mortality). The viruses SARS-CoV1 and SARS-CoV2 recognise ACE2 (Angiotensin-Converting Enzyme 2) receptors, while MERS-CoV recognises the DPP4 (human dipeptidyl-peptidase 4) receptor. Bats are the wild reservoirs for the three coronaviruses that are pathogenic for humans. Intermediate hosts have been found for SARS-CoV1 (palm civet, raccoon dog, badger) and for MERS-CoV (dromedary). For SARS-CoV2, no intermediate host has been identified

after three years of the pandemic. Its origin remains controversial [19].

In Wuhan, the emergence of SARS in 2002 stimulated major research into coronaviruses and, from 2004 onwards, a collection of samples taken from bats captured in caves in southern China and southeast Asia. To date, a stock of almost 15,000 samples (blood, saliva, urine, faeces, etc.) has been built up, making it possible to identify more than 200 SARS-like coronaviruses by RT-PCR and sequencing, of which around a hundred sequences have been published. The reservoir of the SARS-CoV1 virus was identified in 2005 in *Rhinolophus spp* [20]. In 2013, Chinese researchers discovered several coronaviruses in bat samples with *spike* gene sequences very similar to those of SARS-CoV1. These viruses can bind and multiply in HeLa cells expressing the ACE2 receptors from humans, civets or *Rhinolophus affinis* [21]. In a cave in Yunnan Province, the Chinese researchers discovered in 2017 eight unknown bat viruses closely related to SARS-CoV1, including two with a functional Spike and six with RBD deletions. The Spike of one virus (Rs4874) is identical (99.9%) to that of SARS-CoV1 [22].

In 2008, an American team carried out a GoF experiment on a wild-type strain of SARS-like coronavirus (Bat-SCoV). They produced a cDNA that was transfected into cultured cells, obtaining a replicative virus poorly penetrating human cells expressing ACE2. The RBD of Bat-SCoV Spike was then replaced by that of SARS-CoV1. The new virus then penetrated and replicated easily in primate and mouse cells [23]. In 2015, GoF experiments were carried out by Chinese and American teams using the MA15 [24], a SARS-CoV1 virus adapted by passages on BALB/c mice [25]. This MA15 mutant is unable to penetrate human cells. The addition to the MA15 virus of the *spike* gene of a wild-type SARS-like coronavirus (SHC014-CoV), produces a new virus capable of recognising the human ACE2 receptor. This virus was serially transmitted *in vitro* on human respiratory epithelium cells, where it reached titres close to the epidemic strain of SARS-CoV1. *In vivo* passages in human ACE2 transgenic mice, previously described [26], showed significant replication in the lungs. The chimeric virus was then fully synthesised, confirming its high replication and virulence *in vivo*. It was no longer neutralised by anti-SARS-CoV1 antibodies or protected by anti-SARS vaccines [24]. It was concluded that there is a potential risk of re-emergence of SARS-CoV1.

In 2014, it was shown that the MERS-CoV Spike protein has two furin sites that sensitise it to cellular proteases [27]. In 2017, an American team carried out GoF experiments on MERS-CoV. The wild-type virus propagated in human DPP4 receptor-expressing transgenic mice initially causes no disease, but after 30 passages, the virus acquires the property of growing 100 times more in the lungs than the parental virus, leading to lethal infection in mice. Genetic analysis of the MERS viruses after passages shows the acquisition of 13 to 22 mutations, including several in Spike [28]. In 2021, an international team showed that deletion of gene 5 in the MERS-CoV virus led to hypervirulence. This gene normally stimulates interferon production during infection of DPP4 transgenic mice [29]. This shows that a gain-of-function can be obtained by deleting a gene.

Recently in 2023, an American team carried out a Gof experiment on the Omicron variant of SARS-CoV2. This virus carrying numerous mutations in the Spike RBD escapes the humoral immunity of vaccine and is more contagious and less virulent than the original Wuhan-Hu-2 virus from 2020. The Omicron variant causes a moderate non-lethal infection in ACE2-transgenic mice. Replacing the Omicron *spike* gene by that of the Wuhan-Hu-1 (D614G) created a chimeric virus of omicron virus that causes a severe infection in mice with 80% mortality. Further investigation showed that mutating non-structural protein 6 (nsp6) in addition to the S protein was sufficient to recapitulate the attenuated phenotype of Omicron [30].

2. Synthetic biology

In 1965, the first deciphered sequence of a natural polynucleotide was a yeast tRNA-Ala (76 nt) [31], which was then synthesised in 1970 in the form of double-stranded DNA [32]. Chemists then developed solid-phase oligonucleotide synthesis, which enabled the first nucleotidic syntheses by segments of 40 to 80 nt [33]. The first functional gene synthesised by this way was a 207 bp DNA encoding tyrosine suppressor tRNA from *E. coli* [34]. Larger genes were then synthesised by assembling large polynucleotides obtained by ligating

overlapping oligonucleotides [35,36], a tedious and costly process. A 2700 bp plasmid containing the *β-lactamase* gene [37] and the 4917 bp gene encoding the merozoite surface protein (MSP-1) of *P. falciparum* [38] were synthesised. This was followed by considerable advances in nucleotide sequencing and synthesis with the setup of high-throughput platforms, which considerably reduces time and costs [39]. It is now possible to synthesise fragments of 8–30 kb, enough to reconstitute most RNA viruses. Several bacterial genomes have now been chemically synthesised and assembled. In 2008, the 582,970 bp genome of *M. genitalium* was synthesised [40], followed in 2010 by the 1.08 Mb genome of *M. mycoides* [41] and in 2017 the creation of nine strains of *Saccharomyces cerevisiae* in which one or two of the 16 chromosomes have been replaced by synthetic DNA [42]. Finally the complete recoding of the *E. coli* genome was achieved in 2019 [43]. Any microorganism genome can now be synthesised and assembled from the *in silico* sequences.

2.1. *Virus synthesis*

It was possible to produce viruses very early on by enzymatic synthesis from viral genomes, using a DNA polymerase for DNA viruses or a reverse transcriptase for RNA viruses. In 1978, the genome of the RNA phage Qβ (4127 nt) was converted into double-stranded cDNA [44], incorporated into a plasmid and transfected into *E. coli* to produce functional phages. This approach, which subsequently allowed mutations to be introduced into the genomes of RNA viruses, has revolutionised our understanding of their biology [45]. A few years later, the cDNA of a poliovirus inserted into a plasmid was introduced into HeLa cells, enabling the production of infectious polioviruses in very small quantities [46]. Thus, cDNAs from RNA viruses can be easily prepared and used to regenerate either positive-strand (mRNA polarity) or negative-strand RNA viruses, depending upon the virus. This approach was widely used to synthesise numerous RNA viruses, including rabies virus [47], respiratory syncytial virus [48], influenza A virus [49, 50], measles virus [51], Ebola virus [52], bunyavirus (arbovirus) [53] and rotavirus [54].

In 2002, the Spanish flu virus was sequenced using viral sequences obtained by reverse transcriptase from multiple human samples dating back to 1918. The virus was then reconstituted into fragments reassembled into cDNA corresponding to the 8 fragments of the virus. These fragments were transferred into monkey MDCK cells or chicken embryo chorioallantoic cells, previously infected with low virulence H1N1 viruses [55,56]. The viable H1N1 Spanish flu virus was resurrected in this way, as highly virulent viruses [57]. Other RNA viruses have also been reconstituted from viral fragments. One can even reconstitute the genomes of retroviruses that have been part of our chromosomes since time immemorial. In 2006, the entire genome (9472 nt) of a HERV retrovirus was successfully synthesised from endogenous retroviral remnants that had been inserted into human chromosomes for over one million years [58], followed by another infectious HERV-K provirus (designated *Phoenix*) [59]. These reconstituted retroviruses could replicate in cultured human cells. Similarly in 2007, using viral DNA fragments obtained by RT-PCR from chimpanzee faeces, it was possible to reconstitute *de novo* an infectious simian immunodeficiency virus (SIVcpz), which is the virus most closely related to HIV-1 [60]. The genomes of large RNA viruses, such as coronaviruses (29 kb), are difficult to clone and manipulate in *E. coli* due to their size and genome instability. Viral fragments transcribed in cDNA are generated from viral isolates, cloned viral DNA, clinical samples or synthetic DNA. These fragments are then reassembled in a single step and cloned into the artificial chromosome of the yeast *S. cerevisiae*. T7 RNA polymerase was then used to generate the viable virus in yeast. The SARS-CoV2 was thus reconstructed from chemically synthetized clones [61].

Any virus genome can also be synthesised by *de novo* chemical synthesis from the *in silico* sequence. In 2000, the first replicative structure obtained in this way was a hepatitis C virus replicon lacking the structural protein genes [62]. In 2002, the first complete synthesis of a poliovirus RNA virus (7500 bp) was achieved in the absence of a molecular template [63]. The cDNA obtained was transcribed *in vitro* into infectious viral RNA after incubation in HeLa cell extract. At the same time, the complete genome of a DNA virus, phage ΦX 174 (5386 bp), was synthesised in a fortnight by chemical synthesis. The DNA was then transfected into *E. coli*, which produced viable bacteriophages [64].

2.2. *Remodelling entire genomes*

The design of new microorganisms involves engineering and modifying synthetic microbial frames, which is one of the best ways of discovering the fundamental principles of life, leading to improved applications in many fields, including medicine and industry. The principle is inspired by the practice of refactoring used in computing in order to improve existing software. The general aim is to improve the internal structure of an existing system for future use, while maintaining the main functions.

2.2.1. *Genome recoding*

On the scale of the entire genome, the genetic code of a virus or a bacteria can be modified. As a result of the degeneracy of this code, there are several synonymous codons for the same amino acid, some of which are used preferentially in the genomes of different microorganisms. Introducing a codon bias can slow down protein production, due to the use of rare codons. This process has been used to attenuate the proliferation and virulence of polioviruses [65]. This can also be done in a targeted way to restrict the expression of certain genes (de-optimisation), without modifying a single amino acid of proteins encoded by genes [66]. One can also use the bias of codon pairs which are unequally distributed according to the various genomes, with certain highly under-represented, impacting the rate of viral replication [67]. In 2019, it was possible to synthesise and recode the entire *E. coli* genome using 59 codons instead of 64 for the wild-type strain [43].

To improve the performance of a microorganism for industrial purposes, efforts have been made to identify non-essential genes by modifying the microbial chassis. This allows to understand which genes are essential for cellular properties and required to maintain cellular life. In general, a number of computer analyses are first used to define the genes essential for maintaining life. These genes are usually involved in basic metabolism, cell wall metabolism, cell division and DNA metabolism. Through synthetic biology, it is thus possible from natural bacterial or viral genomes to redesign and create new biological entities that do not exist in nature, mainly for medical or industrial applications. It can help to design new viruses by manipulating sequences *in silico*, possibly with the help of Artificial Intelligence.

Because of the widespread use of bench-top synthesizers, this biological synthesis approach needs to be controlled, as it can be a true threat. Engineering and modifying synthetic microbial chassis is one of the best ways not only of discovering the fundamental principles of life, but also of improving applications in the fields of health, medicine, agriculture, veterinary medicine and public health. But it also offers the prospect of malicious use on pathogens.

2.2.2. *Construction of microbial chassis*

There are two strategies for building microbial chassis, top-down or bottom-up [68]. The top-down strategy consists of progressively reducing the microorganism's genome. This process makes it possible to identify the genes that are essential for the survival of the organism by combining computer analysis of the systems, experimental data and models such as those of metabolic, regulatory and signaling networks. A framework is then constructed and modified by inactivating or deleting non-essential genes using various strategies, including the use of DNA-mediated procedures and site-specific recombinases, transposon mutagenesis or CRISPR/Cas system. Examples of this approach include genome reduction in *E. coli* and *B. subtilis*. This has resulted in a faster-growing strain of *E. coli* than the parental strain [69] and elucidation of the genetic basis of secreted proteins in *B. subtilis* [70].

The bottom-up strategy uses low-cost chemical synthesis of DNA segments from various genomes that will be fused, making it possible to construct versatile chassis from different biological entities. The new synthetised DNA fragments were assembled and transplanted into organisms (cells, bacteria, etc.) to create and produce entire genomes. Genetic entities can easily be designed by synthetic biology with the help of computers. For example in 2016, a viable bacterium derived from *M. mycoides* was successfully constructed, harboring a minimal bacterial genome synthetised from an *in silico* sequence of 531 kb instead of the 1079 kb of the wild-type strain [71]. This strategy was also used in 2019 to synthesise a minimal genome of the bacterium *Caulobacter crescentus* [72]. Additionally, in this way, personalised microorganisms can be constructed to produce molecules for industrial use. It is also possible to reformat viruses to improve their performance starting from known genomes. This was done in

2005 with bacteriophage T7, which was physically separated into genomic segments that were then reassembled into an ergonomic virus with improved performance. The chimeric virus is viable with all the properties of the wild-type virus, but is much easier to manipulate [73]. Viral genomes can therefore be easily reconstructed to improve performance. Using the known viral sequence of the horsepox virus [74], researchers succeeded in 2018 in synthesising the horsepox virus (212 kb) by fusing ten large synthesised DNA fragments of 10–30 kb [75]. The resulting DNA was introduced into cells infected with a related poxvirus, the Shope's fibroma virus. Thus, a live horsepoxvirus is produced in cell culture. This virus is less virulent in mice than the vaccinia virus and induces a protective response against it. This is the first complete synthesis of a poxvirus using synthetic biology. This publication is a debatable technological achievement which has been heavily criticised, as it reveals the strategy for constructing in the laboratory a very dangerous virus, the smallpox virus [76, 77].

3. Artificial intelligence

AI's ability to process large quantities of raw, unstructured data (DNA sequences, proteins, etc.) has made it possible to reduce the time and cost of certain experiments, to carry out others that were previously unfeasible and to contribute to the wider field of genetic engineering [78]. One of the applications of AI is the machine learning, which allow to extract knowledge from data and learn from it autonomously. By using algorithms to analyse large quantities of data, this approach enables a machine to learn and improve automatically. We can increase the human capacity to modify genetic material to obtain specific functions. This raises concerns about potential biosafety uses [79].

Applications to biological systems, in particular Machine Learning, enable genetic systems to be modified and programmed with new functions. Engineering principles and the use of systematic design tools are used to reprogram cellular systems. The use of machine learning has led to astonishing progress in the computational design of proteins, enabling industrial and biomedical applications [80]. Synthetic proteins have been designed to carry out cellular functions. The corresponding genes can then be synthesised and inserted into the microbial genome. This can be applied to virulence factors such as toxins or adhesins in pathogenic microorganisms. For example, new toxins can be designed from molecular models, and even dreadful weapons can be created. A recent example is given by Collaborations Pharmaceuticals, which uses an AI system called *MegaSyn*, trained with pesticides, environmental toxins and drugs, in order to find new medicines. Instead of looking for molecules with the lowest possible toxicity, the researchers asked it to look for the highest toxicity and bioactivity. In less than 6 hours, the AI generated more than 40,000 neurotoxic molecules, each more toxic than the last. In particular, it discovered VX, one of the most dangerous nerve agents invented in 1952 and ten times more deadly than sarin. It has also discovered other known chemical weapons and many as yet unknown molecules that are potentially more toxic than VX [81]. The researchers said they had never thought before about how their tool could be hijacked, and were very surprised by the results. With the development of AI, the design of weapons that are still unimaginable today might soon be within everyone's reach.

4. Conclusion

Advances in biotechnology are ushering in a new era, one of unprecedented scientific progress in our knowledge of living organisms, but also one of dangerous biology, for which we must be extremely vigilant. One can now easily modify the genome of microorganisms (loss or gain-of-function) and synthesise gene sequences and even entire microorganisms (bacteria and viruses), enabling major advances in many areas of science. All these manipulations can be aided by the use of AI to design *de novo* manipulations and genome syntheses. These include a better understanding of the evolution and properties of dangerous pathogens. However, this has also implications in terms of dual use and availability of highly dangerous germs that can be synthesised from *in silico* sequences available from databases. In addition, genome synthesis of viruses offers unprecedented possibilities for modifying natural genomes, allowing to create new and potentially dangerous infectious microorganisms. As a result, the Humanity could face the threat of pathogens far worse than anything nature could create. The next big challenge

will be to reconcile scientific progress with biosafety and biosecurity.

Declaration of interests

The author does not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and has declared no affiliations other than his research institution.

References

[1] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, "A programmable Dual-RNA–guided DNA endonuclease in adaptive bacterial immunity", *Science* **337** (2012), p. 816-821.

[2] NCBI Virus, "Viral genomes", Online at https://www.ncbi.nlm.nih.gov/genome/viruses/ (accessed August 9, 2024).

[3] J. Holland, K. Spindler, F. Horodyski, E. Grabau, S. Nichol, S. VandePol, "Rapid evolution of RNA genomes", *Science* **215** (1982), p. 1577-1585.

[4] M. Eigen, "Viral quasispecies", *Sci. Am.* **269** (1993), p. 42-49.

[5] C. Tu, G. Crameri, X. Kong *et al.*, "Antibodies to SARS coronavirus in civets", *Emerg. Infect. Dis.* **10** (2004), p. 2244-2248.

[6] J. H. Ellwanger, J. A. Bogo Chies, "Zoonotic spillover: understanding basic aspects for better prevention zoonotic spillover", *Genet. Mol. Biol.* **44** (2021), no. S1, article no. e20200355.

[7] B. L. Pike, K. E. Saylors, J. N. Fair *et al.*, "The origin and prevention of pandemics", *Clin. Infect. Dis.* **50** (2010), p. 1636-1640.

[8] R. K. Plowright, C. R. Parrish, H. McCallum *et al.*, "Pathways to zoonotic spillover", *Nature Rev. Microbiol.* **15** (2017), p. 502-510.

[9] Z. W. Ye, S. Yuan, K. S. Yuen *et al.*, "Zoonotic origins of human coronaviruses", *Int. J. Bio. Sci.* **16** (2021), p. 1686-1697.

[10] R. J. Jackson, A. J. Ramsay, C. D. Christensen, S. Beaton, D. F. Hall, I. A. Ramshaw, "Expression of mouse interleukin-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox", *J. Virol.* **75** (2001), p. 1205-1210.

[11] L. D. Sims, T. M. Ellis, K. K. Liu *et al.*, "Avian influenza in Hong Kong 1997–2002", *Avian Dis.* **47** (2003), p. 832-838.

[12] K. Ungchusak, P. Auewarakul, S. F. Dowell *et al.*, "Probable person-to-person transmission of avian influenza A (H5N1)", *N. Engl. J. Med.* **352** (2005), p. 333-340.

[13] OMS, "Bulletins d'information sur les flambées épidémiques Grippe aviaire A(H5N1)—Viet Nam", 2 avril 2024. Online at https://www.who.int/fr/emergencies/disease-outbreak-news/item/2024-DON511 (accessed August 9, 2024).

[14] M. Imai, T. Watanabe, M. Hatta *et al.*, "Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets", *Nature* **486** (2012), p. 420-428.

[15] S. Herfst, E. J. A. Schrauwen, M. Linster *et al.*, "Airborne transmission of influenza A/H5N1 virus between ferrets", *Science* **336** (2012), p. 1534-1541.

[16] Y. Zhang, Q. Zhang, H. Kong *et al.*, "H5N1 hybrid viruses bearing 2009/H1N1 virus genes transmit in guinea pigs by respiratory droplet", *Science* **340** (2013), p. 1459-1463.

[17] M. Agüero, I. Monne, A. Sánche *et al.*, "Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022", *Euro. Surveill.* **28** (2023), no. 3, article no. 2300001.

[18] Y. Buisson, "Covid-19, an unfinished story", *Presse. Med. (QMR)* **51** (2022), no. 3, article no. 104131.

[19] P. Berche, "Gain-of-function and origin of Covid19", *Presse. Med. (QMR)* **51** (2022), no. 3, article no. 104167.

[20] W. Li, Z. Shi, M. Yu *et al.*, "Bats are natural reservoirs of SARS-like coronaviruses", *Science* **310** (2005), p. 676-679.

[21] X. Y. Ge, J. L. Li, X. L. Yang *et al.*, "Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor", *Nature* **503** (2013), p. 535-538.

[22] B. Hu, L. P. Zeng, X. L. Yang *et al.*, "Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus", *PLoS Pathog.* **13** (2017), no. 11, article no. e1006698.

[23] M. M. Becker, R. L. Graham, E. F. Donaldson *et al.*, "Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice", *Proc. Natl. Acad. Sci. USA* **105** (2008), p. 19944-19449.

[24] V. D. Menachery, B. L. Yount, K. Debbink *et al.*, "A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence", *Nat. Med.* **21** (2015), p. 1508-1513.

[25] A. Roberts, D. Deming, C. D. Paddock *et al.*, "A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice", *PLoS Pathog.* **3** (2007), no. 1, article no. e5.

[26] P. B. McCray, L. Pewe, C. Wohlford-Lenane *et al.*, "Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus", *J. Virol.* **81** (2007), p. 813-821.

[27] J. K. Millet, G. R. Whittaker, "Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein", *Proc. Natl. Acad. Sci. USA* **111** (2014), p. 15214-15219.

[28] K. Li, C. L. Wohlford-Lenane, R. Channappanavar *et al.*, "Mouse-adapted MERS coronavirus causes lethal lung disease in human DPP4 knockin mice", *Proc. Natl. Acad. Sci. USA* **114** (2017), p. 3119-3128.

[29] J. Gutierrez-Alvarez, L. Wang, K. L. Fernandez-Delgado *et al.*, "Middle east respiratory syndrome coronavirus gene 5 modulates pathogenesis in mice", *J. Virol.* **95** (2021), no. 3, article no. e01172-20.

[30] D.-Y. Chen, C. V. Chin, D. Kenney *et al.*, "Spike and nsp6 are key determinants of SARS-CoV-2 Omicron BA.1 attenuation", *Nature* **615** (2023), p. 143-150.

[31] R. W. Holley, J. Apgar, G. A. Everett *et al.*, "Structure of a ribonucleic acid", *Science* **147** (1965), p. 1462-1465.

[32] K. L. Agarwal, H. Büchi, M. H. Caruthers *et al.*, "Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast", *Nature* **227** (1971), p. 27-34.

[33] M. H. Caruthers, "Gene synthesis machines: DNA chemistry and its uses", *Science* **230** (1985), p. 281-285.

[34] H. G. Khorana, K. L. Agarwal, P. Besmer *et al.*, "Total synthesis

of the structural gene for the precursor of a tyrosine suppressor transfer RNA from *Escherichia coli*. 1. General introduction", *J. Biol. Chem.* **251** (1978), p. 565-570.

[35] L. Stewart, A. B. Burgin, "Whole gene synthesis: a gene-o-matic future", *Front. Drug Des. Disc.* **1** (2005), p. 297-341.

[36] Y. Sanghvi, "A roadmap to the assembly of synthetic DNA from raw materials", in *Working Papers for Synthetic Genomics: Risks and Benefits for Science and Society* (M. S. Garfinkel, D. Endy, G. L. Epstein, R. M. Friedman, eds.), Craig Venter Institute, Rockville, Md, 2007, p. 17-33.

[37] W. P. Stemmer, A. Crameri, K. D. Ha *et al.*, "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides", *Gene* **164** (1995), p. 49-53.

[38] W. Pan, E. Ravot, R. Tolle *et al.*, "Vaccine candidate MSP-1 from *Plasmodium falciparum*: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from *E. coli* and mammalian cells", *Nucleic Acids Res.* **27** (1999), p. 1094-1103.

[39] J. A. Shendure, G. J. Porreca, G. M. Church, "Overview of DNA sequencing strategies", in *Current Protocols in Molecular Biology* (F. M. Ausubel *et al.*, eds.), John Wiley & Sons, Hoboken, NJ, 2008.

[40] D. G. Gibson, E. Ravot, R. Tolle *et al.*, "Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome", *Science* **319** (2008), p. 1215-1220.

[41] D. G. Gibson, J. I. Glass, C. Lartigue *et al.*, "Creation of a bacterial cell controlled by a chemically synthesized genome", *Science* **329** (2010), p. 52-56.

[42] S. M. Richardson, L. A. Mitchell, G. Stracquadanio *et al.*, "Design of a synthetic yeast genome", *Science* **355** (2017), p. 1040-1044.

[43] J. Fredens, K. Wang, D. de la Torre *et al.*, "Total synthesis of *Escherichia coli* with a recoded genome", *Nature* **569** (2019), p. 514-518.

[44] T. Taniguchi, M. Palmieri, C. Weissmann, "Qβ DNA-containing hybrid plasmids giving rise to Qβ phage formation in the bacterial host", *Nature* **274** (1978), p. 223-228.

[45] C. Weissmann, H. Weber, T. Taniguchi *et al.*, "Reversed genetics: a new approach to the elucidation of structure–function relationship", *Ciba. Found. Symp.* **66** (1979), p. 47-61.

[46] V. R. Racaniello, D. Baltimore, "Cloned poliovirus complementary DNA is infectious in mammalian cells", *Science* **214** (1981), p. 916-919.

[47] M. J. Schnell, T. Mebatsion, K. K. Conzelmann, "Infectious rabies viruses from cloned cDNA", *EMBO J.* **13** (1994), p. 4195-4203.

[48] P. L. Collins, M. G. Hill, E. Camargo *et al.*, "Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5′ proximal open reading frame of the M2 mRNA in gene expression and provides a capability for vaccine development", *Proc. Natl. Acad. Sci. USA* **92** (1995), p. 11563-11567.

[49] G. Neumann, T. Watanabe, H. Ito *et al.*, "Generation of influenza A viruses entirely from cloned cDNAs", *Proc. Natl. Acad. Sci. USA* **96** (1999), p. 9345-9350.

[50] E. Fodor, L. Devenish, O. G. Engelhardt, P. Palese, G. G. Brownlee, A. García-Sastre, "Rescue of influenza a virus from recombinant DNA", *J. Virol.* **73** (1999), p. 9679-9682.

[51] K. Takeuchi, M. Takeda, N. Miyajima, "Toward understanding the pathogenicity of wild-type measles virus by reverse genetics", *Jpn. J. Infect. Dis.* **55** (2002), p. 143-149.

[52] G. Neumann, H. Feldmann, S. Watanabe *et al.*, "Reverse genetics demonstrates that proteolytic processing of the Ebola virus glycoprotein is not essential for replication in cell culture", *J. Virol.* **76** (2002), p. 406-410.

[53] A. K. Överby, V. Popov, E. P. Neve, R. F. Pettersson, "Generation and analysis of infectious virus-like particles of uukuniemi virus (bunyaviridae): a useful system for studying bunyaviral packaging and budding", *J. Virol.* **80** (2006), p. 10428-10435.

[54] S. Komoto, J. Sasaki, K. Taniguchi, "Reverse genetics system for introduction of site-specific mutations into the double-stranded RNA genome of infectious rotavirus", *Proc. Natl. Acad. Sci. USA* **103** (2006), p. 4646-4651.

[55] A. H. Reid, T. G. Fanning, T. A. Janczewski *et al.*, "Characterization of the 1918 Spanish influenza virus matrix gene segment", *J. Virol.* **76** (2002), p. 10717-10723.

[56] J. K. Taubenberger, A. H. Reid, R. M. Lourens *et al.*, "Characterization of the 1918 influenza virus polymerase genes", *Nature* **437** (2005), p. 889-893.

[57] T. M. Tumpey, C. F. Basler, P. V. Aguilar *et al.*, "Characterization of the reconstructed 1918 Spanish influenza pandemic virus", *Science* **310** (2005), p. 77-80.

[58] Y. N. Lee, P. D. Bieniasz, "Reconstitution of an infectious human endogenous retrovirus", *PLoS Pathog.* **3** (2007), article no. e10.

[59] M. Dewannieux, F. Harper, A. Richaud *et al.*, "Identification of an infectious progenitor for the multiplecopy HERV-K human endogenous retroelements", *Genome Res.* **16** (2006), p. 1548-1556.

[60] J. Takehisa, M. H. Kraus, J. M. Decker *et al.*, "Generation of infectious molecular clones of simian immunodeficiency virus from fecal consensus sequences of wild chimpanzees", *J. Virol.* **81** (2007), p. 7463-7475.

[61] T. T. N. Thao, F. Labroussaa, N. Ebert *et al.*, "Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform", *Nature* **582** (2020), p. 561-565.

[62] K. J. Blight, A. A. Kolykhalov, C. M. Rice, "Efficient initiation of HCV RNA replication in cell culture", *Science* **290** (2000), p. 1972-1974.

[63] J. Cello, A. V. Paul, E. Wimmer, "Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template", *Science* **297** (2002), p. 1016-1018.

[64] H. O. Smith, C. A. Hutchison, C. Pfannkoch, J. C. Venter, "Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides", *Proc. Natl. Acad. Sci. USA* **100** (2003), p. 15440-15445.

[65] C. C. Burns, J. Shaw, R. Campagnoli *et al.*, "Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region", *J. Virol.* **80** (2006), p. 3259-3272.

[66] S. Mueller, D. Papamichail, J. R. Coleman *et al.*, "Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity", *J. Virol.* **80** (2006), p. 9687-9696.

[67] J. R. Coleman, D. Papamichail, S. Skiena *et al.*, "Virus atten-

uation by genome-scale changes in codon pair bias", *Science* **320** (2008), p. 1784-1787.

[68] H. Chia, X. Wanga, Y. Shaoa *et al.*, "Engineering and modification of microbial chassis for systems and synthetic biology", *Synth. Syst. Biotechnol.* **4** (2019), p. 25-33.

[69] H. Mizoguchi, H. Mori, T. Fujio, " *Escherichia coli* minimum genome factory", *Biotechnol. Appl. Biochem.* **46** (2007), p. 157-167.

[70] H. Tjalsma, H. Antelmann, J. D. Jongbloed *et al.*, "Proteomics of protein secretion by *Bacillus subtilis*: separating the 'secrets' of the secretome", *Microbiol. Mol. Biol. Rev.* **68** (2004), p. 207-233.

[71] C. A. Hutchison, R. Y. Chuang, V. N. Noskov *et al.*, "Design and synthesis of a minimal bacterial genome", *Science* **351** (2016), no. 6280, article no. aad6253.

[72] J. E. Venetz, L. Del Medico, A. Wölfle *et al.*, "Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality", *Proc. Natl. Acad. Sci. USA* **116** (2019), p. 8070-8079.

[73] L. Y. Chan, S. Kosuri, D. Endy, "Refactoring bacteriophage T7", *Mol. Syst. Biol.* **1** (2005), article no. 2005.0018.

[74] E. R. Tulman, G. Delhon, C. L. Afonso *et al.*, "Genome of horsepox virus", *J. Virol.* **80** (2006), p. 9244-9258.

[75] R. S. Noyce, S. Lederman, D. H. Evans, "Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments", *PLOS One* **13** (2018), article no. e0188453.

[76] G. D. Koblentz, "The *de novo* synthesis of horsepox virus: implications for biosecurity and recommendations for preventing the reemergence of smallpox", *Health Secur.* **15** (2017), no. 6, p. 620-628.

[77] T. Inglesby, "Horsepox and the need for a new norm, more transparency, and stronger oversight for experiments that pose pandemic risks", *PLOS Pathog.* **14** (2018), article no. e1007129.

[78] A. Bhardwaj, S. Kishore, D. K. Pandey, "Artificial intelligence in biological sciences", *Life* **12** (2022), article no. 1430.

[79] T. Kuiken, "Artificial intelligence in the biological sciences: uses, safety, security, and oversight, congressional reseach service", R47949, Washington, Nov. 22, 2023. Online at https://crsreports.congress.gov/product/pdf/R/R47849 (accessed August 9, 2024).

[80] S. Kohyama, B. P. Frohn, L. Babl, P. Schwille, "Machine learning-aided design and screening of an emergent protein function in synthetic cells", *Nat. Commun.* **15** (2024), article no. 2010.

[81] F. Urbina, F. Lentzos, C. Invernizzi, S. Ekins, "Dual use of artificial intelligence-powered drug discovery", *Nat. Mach. Intell.* **4** (2022), no. 3, p. 189-191.