Meeting of the States Parties to the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction

11 June 2012

Original: English

2012 Meeting
Geneva, 10–14 December 2012

Meeting of Experts
Geneva, 16–20 July 2012
Item 6 of the provisional agenda
Standing agenda item: review of developments in the field of science and technology related to the Convention

Advances in enabling technologies

Background information document submitted by the Implementation Support Unit

Summary

The Seventh Review Conference decided that the 2012 to 2015 intersessional programme would include a Standing Agenda Item on review of developments in the field of science and technology related to the Convention. The Conference also decided that under this item in 2012, States Parties would consider "advances in enabling technologies, including high-throughput systems for sequencing, synthesizing and analyzing DNA; bioinformatics and computational tools; and systems biology". This paper provides an overview of advances of possible relevance. It expands upon and updates the background information document on new scientific and technological developments relevant to the Convention prepared for the Seventh Review Conference (BWC/CONF.VII/INF.3 and addenda). The annex, in English only, provides a more detailed account with references to the scientific literature.

I. Characterizing biological systems and networks

1. Considerable progress has been made in recent years across a broad range of different "-omics", such as genomics (the study of all the genetic information in an organism), transcriptomics (the study of all the RNA in an organism), proteomics (the study of all the proteins in an organism), metabolomics (the study of all the biochemical processes or metabolism of an organism), as well as how they relate to one another.

2. Genomics advances have included: genome wide analysis; progress in understanding the role of single nucleotide polymorphisms (SNPs) in disease; progress in understanding

the role of copy number variation in disease; functional genomics; and increased understanding of the evolvability of gene regulatory networks.

3. Advances in transcriptomics include: the identification of regulators; the characterization of regulators; and the implications of network structure.

4. Progress in proteomics has included: better understanding of how proteins are synthesized; fluctuations in their presence over time; better characterization of the system which ensures the premature termination of sequences that fail quality control; new tools to assist in the identification and quantification of proteins; increasing standardization of data reporting; improved tools for determining the structure of proteins; enhanced understandings of protein-protein interactions, such as through mapping, regulation, cross network comparisons and studying protein signalling cascades.

5. Metabolimics advances have included: comparative studies of pathways between species; improved tools for perturbing and studying pathways; investigations of network motifs; as well as studies on fluxes within metabolic networks (fluxomics).

6. There has also been considerable progress made in integrating data from these fields, especially in terms of mapping and, to a lesser extent, modelling systems. Perhaps the best example of combining different approaches was the characterization of *Mycoplasma pneumoniae* which included the integration of genomic, metabolic, proteomic, structural and cellular information.

II. Manipulating biological systems and networks

7. There have been a variety of developments over the last five years that enable greater control in manipulating biological systems and networks. The two most significant advances were RNA interference technology (RNAi) and Zinc Finger Nucleases (ZFN).

III. Engineering biological systems and networks

8. Biological engineering, or synthetic biology, has advanced considerably over the past five years. Industry is becoming increasingly interested in these approaches. There has been a significant increase in the biological complexity of the biological systems and networks that can be engineered.

9. In addition to the chemical synthesis of a genome able to control a bacterial cell (Craig Venter's artificial life) other important stepping stones include: the engineering of the metabolic pathway in yeast to produce the precursor of an anti-malarial drug; the creation of a synthetic mammalian gene circuit that revealed anti-tuberculosis compounds; a demonstration of distributed biological computation; and the engineering of an *E. coli* to sense and kill a human pathogen.

10. There has been progress in overcoming the technical hurdles identified as limiting the utility of synthetic biology, including: characterizing parts; improving wiring; addressing complexity; improved interoperability; and enhanced reliability. There have been: technical improvements; chassis improvement; and the development of new components. Significant attention has also been paid to the safety and security implications of these advances.

11. Biomedical applications are beginning to emerge, including for: understanding disease mechanisms; disease prevention; drug development; novel treatment for infections; and cancer therapies.

IV. Gathering and manipulating biological information

12. Advances in bioinformatics and computational biology have greatly aided the gathering, processing and utility of biological data, including: the creation of new languages; advances in data mining; improvements in modelling and simulation, including the creation of whole-cell simulations; online tools and software for visualising complex biological information, analysing gene sequence data, protein analysis; as well as designing tools. Laboratories are becoming increasingly digitised. Advances in bioinformatics have been combined with progress in characterization technology, high-throughput approaches and robotics to create a fully automated researcher. A computer controlled artificial intelligence develops hypotheses, tests them in an automated laboratory and feeds the results back into the system, to design a new round of experiments. Not only do robot scientists promise to take much of the drudgery out of basic research but they might also help to address the current bottlenecks in characterizing parts, identifying function and interpreting raw data.

V. Converting biological information to digital data and back

13. If biology is becoming an information science then in part it is because of the ability to convert biological data into digital data and back again. Gene sequencing (reading the genetic code) enables researchers to move in one direction and gene synthesis (writing the genetic code) the other. Capabilities to read and write genetic code are not new but capabilities in these areas have changed dramatically over the past five years.

14. Over the past five years second and then third generation sequencers have become available. This has led to a dramatic increase in raw sequencing power. Modern machines can sequence a human genome in a day. The cost of sequencing a human genome has fallen to under $1000. This has enabled new types of project to be attempted and different types of data collected. At the time of the Sixth Review Conference in 2006, only two human genomes had been sequenced. As of October 2011, over 13,000 human genomes had been sequenced.

15. Novel health applications are being found for this expanded sequencing capacity including in diagnosis and guiding therapy. Both governments and the private sector have invested heavily in developing new applications, tools and platforms.

16. Trends in synthesis capacity mirror those for sequencing. There have been technical improvements in the ability to produce longer strands of genetic material. New assembly techniques make it easier and faster to combine short fragments into long sequences. The cost of having gene length fragments commercially synthesized also continues to fall. The quality of sequenced material seems to be increasing. This has led to more sophisticated projects being attempted. Over the past five years, synthesis of genetic material has moved from viral settings, through bacterial settings, and mammalian organelles, to partial synthesis of a chromosome from a eukaryote.

VI. Generic enabling technologies

17. Underpinning many of the advances discussed throughout this paper are a range of technologies that make it easier, cheaper, faster or more reliable to do many of the basic procedures and practices involved in expanding the limits of current understanding and creating new applications. Other advances have allowed scientists to do things that were previously unattainable. There has been a broad range of new enabling technologies developed over the past five years.

Annex

[ENGLISH ONLY]

Advances in enabling technologies: a more detailed review

I. Characterizing biological systems and networks

1. Considerable progress has been made in recent years across a broad range of different "-omics", such as genomics (the study of all the genetic information in an organism), transcriptomics (the study of all the RNA in an organism), proteomics (the study of all the proteins in an organism), metabolomics (the study of all the biochemical processes or metabolism of an organism), as well as how they relate to one and other.

2. Genomic advances have included: a deeper understanding of the importance of "junk" genetic material[1]; a more sophisticated appreciation of how and why genes are expressed, through epigenomics[2]; developments in identifying genetic interactions, especially through the use of RNAi;[3] a better understanding of impact of mutations in hotspots, (or quantitative trait loci) on the downstream expression of distant genes;[4] and new techniques to identify novel or rare genomes from collected genomic data.[5] Advances related to genome wide analysis have:[6] enabled the simultaneous analysis of single nucleotide polymorphisms (SNPs) to identify higher level interactions;[7] led to efforts to understand how SNPs relate to disease; provided new insights in transcription;[8] as well as provided insights into the genetic component of social behaviour.[9] One example of a study that has linked genomics to disease was the investigation of SNP variation in the genomic epidemiology of malaria.[10] Parallel progress in the implications of copy number variations include: their role in gene and genome evolution; their impact on gene expression profiles; as well as their relationship with disease.[11] There have also been advances in functional genomics,[12] such as: creating a genome wide functional map of genes in a mammal; using evolutional developmental biology to help bridge the gap between genetic information and physical characteristic; and in using RNAi to understand epistatic genetic interactions.[13] There has also been considerable development of concepts of the evolvability of gene regulatory networks.[14] Research has shown, for example, how gene networks develop

[1] http://www.newscientist.com/article/dn14667-junk-dna-may-have-handed-us-a-gripping-future.html
[2] http://www.nature.com/news/2010/100510/full465145a.html
[3] http://www.nature.com/nmeth/journal/v8/n4/full/nmeth.1581.html
[4] http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000232
[5] http://www.sciencemag.org/content/335/6068/587.abstract
[6] http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1000218
[7] http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000130
[8] http://www.nature.com/nature/journal/v483/n7389/abs/nature10799.html
[9] http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000127
[10] http://www.nature.com/nature/journal/v456/n7223/full/nature07632.html
[11] http://www.annualreviews.org/doi/abs/10.1146/annurev.genom.9.081307.164217
[12] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000165
[13] http://www.nature.com/nmeth/journal/v8/n4/full/nmeth0411-299.html
[14] http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1000112

robustness through the application of selective pressures, such as provided by host-parasite interactions.[15]

3. Advances in transcriptomics can be roughly broken down into the identification of regulators, the characterization of regulators, and those that relate to network structure.[16] Studies released over the past five years have identified a number of transcriptomic regulators, such as microRNAs (miRNAs), piwi-interacting RNAs, and small interfering RNA (siRNA).[17] Our understanding of the roles played by such regulators has also expanded including: in explaining the comparative complexity of different organisms; in regulating gene expression; in evolutionary development; and in determining the phenotypic (physical) properties of plants. Progress has been made in characterizing regulators, including a quantitative comparison of the short RNA-based systems and protein-based gene regulation.[18] There has also been an advance in our understanding of the role of large intergenic non-coding RNAs (lincRNAs) which have been shown to regulate gene expression. Studies of the control networks for transcription have highlighted that their topography has implications for function.[19] They seem to be organised to avoid malfunctions. Their robustness also seems to be linked to their structure, specifically the volume and geometry of flexible regions in the parameter space.[20]

4. Considerable progress has been made across the field of proteomics. Understanding of how proteins are synthesised, for example, has been supplemented by better characterization of the system which ensures the premature termination of sequences that fail quality control.[21] Other advances have helped explain how protein composition changes over time, for example, through insights into the structure and function of enzymes responsible for their degradation.[22] There have been new tools assist in the identification and quantification of proteins,[23] such as: electron-vibration-vibration two-dimensional infrared spectroscopy; and advances in mass spectrometry. Guidelines have also been developed for facilitate the standardization of data reporting in proteomics, including for mass spectrometry and gel electrophoresis. In terms of determining the structure of proteins, there have been a series of advances in developing high-throughput approaches,[24] including in detecting mature and changing forms of proteins.[25] Similar advances have enabled the structures of "once-intractable" proteins to be identified.[26] Structural comparisons of proteins in different species have also enabled researchers to make headway in determining the function of specific proteins.[27] Perhaps the area of greatest interest has been in working on protein-protein interactions (PPI) with progress being made in mapping, regulation, cross network comparisons and protein signalling cascades. [28]

5. PPI maps have been generated using high-throughput microfluidic approaches. Additional details have been added from studying mRNAs.[29] These maps have improved

[15] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2516366/
[16] http://www.nature.com/nature/journal/v455/n7217/full/4551184a.html
[17] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2583084/
[18] http://www.nature.com/nature/journal/vaop/ncurrent/full/nature10398.html
[19] http://phys.org/news192128818.html
[20] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000256
[21] http://www.nature.com/nature/journal/v457/n7226/full/457157a.html
[22] http://www.nature.com/nature/journal/vaop/ncurrent/full/nature10774.html
[23] http://www.nature.com/nmeth/journal/v5/n12/full/nmeth1208-993.html
[24] http://www.sciencemag.org/site/products/lst_20080801.xhtml
[25] http://www.nature.com/nature/journal/v480/n7376/full/nature10575.html
[26] http://www.nature.com/news/opioid-receptors-revealed-1.10273
[27] http://www.biomedcentral.com/1752-0509/2/69
[28] http://www.ncbi.nlm.nih.gov/pubmed/19098921
[29] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2387231/

our understanding of cellular organization and function.[30] They could also act as an important resource for annotating the proteome.[31] Considerable effort has gone into refining the topology of maps, including the roles of: hubs; and randomness.[32] [33] The importance of including structural information in the maps, for example, has been demonstrated.[34] The regulation of PPI has led to improvements in our understanding of how protein complexes form.[35] The constraints placed upon PPIs by non-functioning interactions have also been investigated.[36] Research released over the past five years links the regulation of PPI to innate immunity.[37] By studying protein interaction networks in different organisms, researchers have been able to identify conserved protein function.[38] Published results also highlight recurring design patterns in network design.[39] There are also shared mechanisms within the various network schemas.[40] There have also been a range of advances relating to the characterization of protein signalling cascades. One group examined dynamic capabilities and used the results to help them identify functions.[41] A second group both quantified information exchange and determined channel noise and capacity.[42] Insights into the regulation of protein signalling cascades have come from investigating the roles of signal duration.[43]

6. The field of metabolomics is evolving from "cataloguing metabolites to asking broader biological questions about how metabolites reflect and affect cell function".[44] For example, comparing metabolic pathways between species provides information on their evolution, can assist in metabolic engineering and may assist in analysing diseases and designing drugs.[45] There have been advances in the tools available to study metabolomics, including allowing the targeting of simultaneous perturbations to determine the structure and function of networks.[46] The study of certain network motifs has facilitated determination of how and when certain pathways within networks are used.[47] Research has also indicated that fluxes within metabolic networks (the study of which is sometimes called fluxomics) are connected to health and disease.[48] The related field of studying "the global, dynamic metabolic response of living systems to biological stimuli or genetic manipulation" (metabonomics) has the potential to offer insights into disease networks and assist in drug discovery. [49]

7. Some of the most insightful advances have resulted when data from two or more of these approaches has been combined. For example, structure network analysis has provided

[30] http://www.ncbi.nlm.nih.gov/pubmed/18949022
[31] http://www.ncbi.nlm.nih.gov/pubmed/16169070
[32] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000114
[33] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000140
[34] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2290937/
[35] http://www.biomedcentral.com/1752-0509/3/3
[36] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2538908/
[37] http://genomebiology.com/2008/9/8/R123
[38] http://www.pnas.org/content/105/35/12763.full
[39] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2424294/
[40] http://www.ncbi.nlm.nih.gov/pubmed/18949022
[41] http://www.nature.com/nchembio/journal/v4/n11/full/nchembio1108-643.html
[42] http://www.ncbi.nlm.nih.gov/pubmed/19149897
[43] http://www.biomedcentral.com/1752-0509/2/108
[44] http://www.nature.com/nmeth/journal/v8/n2/full/nmeth0211-117.html
[45] http://www.biomedcentral.com/1752-0509/2/111
[46] http://www.nature.com/nbt/journal/v27/n2/full/nbt0209-149.html
[47] http://www.nature.com/nbt/journal/v26/n11/abs/nbt.1499.html
[48] http://www.nature.com/nbt/journal/v26/n10/full/nbt1008-1090.html
[49] http://www.nature.com/nature/journal/v455/n7216/full/4551054a.html

insights into protein-DNA interactions.[50] Graph alignment of protein and genetic information has provided for additional functional data in at least one pathogen.[51] Another study that made use of protein-DNA interactions produced models for the feedback control of single genes and pairs of genes (toggle switches).[52] Additionally, studies that combined both metabolomic and proteomic data have demonstrated that the relationship between the two can be asymmetrical.[53] A second group used similar sets of data to identify novel molecular organizing principles. [54]

8. There have been significant advances in mapping and modelling networks based upon mixed data sets. One map of a cancer-causing pathway, for example, included information on proteins, genes, protein complexes, chemical compounds and biochemical reactions.[55] Creating these maps allows for the identification of higher-order combination effects (where contributing components are found in different approaches).[56] Mapping efforts have also begun to evolve into modelling attempts. One group reported developing a genome-scale kinetic model which combines genomic data with metabolic data and fluxomic data.[57]

9. Perhaps one of the most impressive examples of what can be achieved through combining these different approaches was the characterization of *Mycoplasma pneumoniae* which included the integration of genomic, metabolic, proteomic, structural and cellular information.[58] Combining -omics can also provide direct insights into disease. There have been efforts to reverse engineer the networks responsible for complex diseases.[59] Researchers have also reported the development of a computational framework that integrates proteomic information, similarities in disease phenotype and known gene-phenotype associations to assist in identifying currently unknown disease-related genes.[60]

II. Manipulating biological systems and networks

10. There have been a variety of developments over the last five years that enable greater control in manipulating biological systems and networks. Researchers have proven successful in unlocking capacity in such systems, for example, by reactivating latent viruses.[61] There have also been practical advances in sidestepping interruptions in metabolic networks — either by bypassing the affected genes or by compensating for functions via network manipulation.[62] Researchers have also developed our abilities to manipulate the growth rates of cellular cultures[63]and to manipulate muscle mass and exercise endurance in animals.[64] There have also been significant developments in ability to

[50] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000170
[51] http://www.biomedcentral.com/1752-0509/2/90
[52] http://www.biomedcentral.com/1752-0509/2/94
[53] http://genomebiology.com/content/10/2/R19
[54] http://www.biomedcentral.com/1752-0509/2/100
[55] http://www.ncbi.nlm.nih.gov/pubmed/18319725
[56] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2538911/
[57] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2290940/
[58] http://www.embl.de/aboutus/communication_outreach/media_relations/2009/091127_Heidelberg/
[59] http://www.biomedcentral.com/1752-0509/2/72
[60] http://www.nature.com/msb/journal/v4/n1/full/msb200827.html
[61] http://online.wsj.com/article/SB10001424052748704529204576256714090044534.html
[62] http://www.nature.com/msb/journal/v4/n1/full/msb20081.html
[63] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000257
[64] http://www.cell.com/abstract/S0092-8674%2811%2901223-2

engineer controls for networks.[65] One group has reported rewiring RNA machinery to include an on/off switch that can be manipulated by the addition of endogenous proteins. The proof-of-principle has been built into human T-cells.[66] A second team engineered a light-activated on-off switch for use in animal models.[67] The discovery of novel inter-cellular communication channels in bacteria also offers additional routes to add information into, or take it out of these systems.[68] The two most significant advances in this area, however, have been with RNAi technology and Zinc Finger Nucleases (ZFN).

11. RNAi is a mechanism which silences individual genes. It is used in nature as one of the many small RNAs that regulates transcription. It is a powerful research tool as it enables the direct perturbation of the genetic network by being programmed to silence virtually any genetic sequence. Over the past five years considerable progress has been made in understanding its biochemical and biophysical properties and describing the various mechanism by which it works.[69] RNAi has been used in public health research, for example, to examine how drugs to treat African sleeping sickness enter cells and exert biological effects.[70] There have also been advances that facilitate more programmable control over RNAi, especially through model-guided design.[71] There has been considerable interest in the therapeutic potential of RNAi.[72] For example, the World Organization for Animal Health has highlighted its potential for combating foot-and-mouth disease and in interfering with influenza infections in poultry. Recent years have seen large pharmaceutical companies turning away from developing RNAi-based therapies.[73] Smaller companies are making progress in developing RNAi-based products.[74] Studies of patent applications, and patents granted, however, suggest that there is still significant commercial interest in this technology.[75] One of the technical challenges to developing RNAi-based therapeutics has been getting it inside cells. In July 2011, a research team reported have created a new nanoparticle-based delivery system that might overcome this hurdle.[76]

12. ZFNs are a powerful genome engineering tool which can be targeted to a particular genomic domain, cuts both strands of the DNA and allows for donor DNA to be added instead.[77] This enables both gene deletion and site-specific mutations. ZFNs have been used to delete up to 15 million bases of information. Until very recently they have been difficult to design and produce. It has been a task left to specialist contractors.[78] Three papers published in early 2011 report: more streamlined production via context-dependent assembly (CoDA) which might open doors to in house production of ZFN;[79] the re-engineering of the dimerization interface creating higher levels of cleavage activity; and improved modular assembly techniques.[80] These papers could open the door for the much

[65] http://www.nature.com/nmeth/journal/v8/n2/full/nmeth0211-108a.html
[66] http://www.technologyreview.com/biomedicine/25237/
[67] http://dev.biologists.org/content/139/9/1691
[68] http://www.cell.com/abstract/S0092-8674(11)00016-X
[69] http://www.jbc.org/content/284/27/17897
[70] http://www.nature.com/nature/journal/vaop/ncurrent/full/nature10771.html
[71] http://www.nature.com/msb/journal/v4/n1/full/msb200862.html
[72] http://www.oie.int/doc/document.php?numrec=3638903
[73] http://www.nature.com/news/2011/110803/full/476010a.html
[74] http://www.genengnews.com/gen-articles/use-of-sirna-in-therapeutic-arena-on-the-upswing/4072/
[75] http://www.nature.com/nbt/journal/v29/n6/full/nbt.1885.html
[76] http://www.masshightech.com/stories/2011/07/25/daily10-Alnylam-and-MIT-publish-RNAi-nanoparticle-findings.html
[77] http://www.nature.com/nmeth/journal/v8/n1/full/nmeth.f.328.html
[78] http://www.nature.com/nmeth/journal/v8/n1/full/nmeth.1542.html
[79] http://www.nature.com/nmeth/journal/v8/n1/full/nmeth0111-53.html
[80] http://www.nature.com/nmeth/journal/v8/n1/full/nmeth.1539.html

wider use of this technology. One stumbling block yet to be overcome is the patent estate associated with this technology. One company now controls the majority of associated intellectual property.[81] Whilst this will likely impact upon opportunities for the commercial development of any discovery made with this system, it is unclear what the implications might be for its use as a research tool. [82]

III. Engineering biological systems and networks

13. Biological engineering, or synthetic biology, has advanced considerably over the past five years. Industry is becoming increasingly interested in these approaches. Synthetic biology has evolved from a field with a great deal of potential, to an approach that is already yielding concrete examples of its potential power (but still with a great deal of potential to grow further). In addition to the chemical synthesis of a genome able to control a bacterial cell (Craig Venter's artificial life) other important stepping stones include: the engineering of the metabolic pathway in yeast to produce the precursor of an anti-malarial drug;[83] the creation of a synthetic mammalian gene circuit that revealed anti-tuberculosis compounds;[84] a demonstration of distributed biological computation; and the engineering of an *E. coli* to sense and kill a human pathogen.[85]

14. The complexity of what can be accomplished using synthetic biology has been increasing.[86] Traditional genetic engineering approaches, which involved the engineering of single genes, were supplemented by metabolic pathway engineering, such as new modular circuits for gene transcription or engineer *E. coli* to produce putrescine.[87] Metabolic pathway engineering was supplemented by the ability to engineer entire organisms, for example engineering *E. coli* to be able to solve mathematical puzzles like the Burnt Pancake Problem or the Hamilton Path Problem.[88,89] Benign viruses have been re-engineered into assembly devices.[90] More recently, the ability to engineer individual organisms has been supplemented with capabilities to engineer collectives of organisms,[91] for example to synchronize blinking patterns or to model a predator-prey ecosystem.[92] Subsequent research has significantly increased the size of colony which can be controlled[93] and the complexity of the behaviour which can be encoded.[94]

15. There are still hurdles to be overcome if biological engineering is going to live up to its full potential. In January 2010 an article in Nature set out five grand challenges:

(a) Many of the parts continue to be uncharacterized;

(b) The 'wiring' of biological circuits remains unpredictable;

(c) The complexity of systems make them difficult to manipulate;

[81] http://www.nature.com/nbt/journal/v27/n2/abs/nbt0209-140.html
[82] http://www.nature.com/nmeth/journal/v8/n1/full/nmeth0111-7a.html
[83] http://www.sciencemag.org/content/329/5987/52.abstract
[84] http://www.pnas.org/content/105/29/9994.abstract
[85] http://www.nature.com/msb/journal/v7/n1/full/msb201155.html
[86] http://www.jbioleng.org/content/4/1/14
[87] http://www.ncbi.nlm.nih.gov/pubmed/19714672
[88] http://www.jbioleng.org/content/2/1/8
[89] http://www.jbioleng.org/content/2/1/8
[90] http://www.nature.com/nature/journal/v478/n7369/abs/nature10513.html
[91] http://www.ncbi.nlm.nih.gov/pubmed/18414488
[92] http://www.sciencemag.org/content/333/6047/1315
[93] http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature10722.html
[94] http://www.pnas.org/content/early/2011/09/19/1109554108.abstract

(d) Many of the parts do not work together as expected; and

(e) Circuits tend not to be reliable thanks to variability.[95]

There has been progress in addressing these challenges. The development of standards for characterization will help to address undefined parts — although there is a great deal of laboratory work to be done on implementing this.[96] Efforts to address the wiring challenge have included: efforts to improve the separation of signal from noise;[97] efforts to reduce biological noise;[98] efforts to work with biological noise;[99] efforts to produce noise-tolerant and delay-robust gene circuits;[100] as well as efforts to incorporate distributed robustness.[101] Improvements in identifying and defining modularity will help to address the levels of complexity involved.[102] Research has also demonstrated that the basic principles of a bottom-up approach to biological engineering work with sufficient modelling and characterization.[103] This suggests that as capabilities in these areas increase, issues of the incompatibility of parts might decrease. Reliability issues are slowly being addressed by improvements in designing evolutionary robust gene circuits and in stabilizing synthetic data in the DNA of living organisms.[104,105]

16. Over the past five years, there have also been advances in: the protocols available for synthetic biology, such as improvements in how synthetic gene circuits can be assembled and optimised;[106] design tools, such as the creation of a computer-aided design tool for synthetic biology:[107] as well as the availability of parts,[108] in terms of the creation of professional facilities to produce parts, developments in the intellectual property frameworks that govern use of those parts, and calls for the publication of full sequence data for synthetic sequences, facilitating the recreation of parts.[109]

17. There have also been advances in the chassis developed for use in synthetic biology.[110] The potential for host physiology to modulate engineered gene circuits highlights the importance of developing efficient chassis. Mechanisms to insulate engineered metabolic circuits from host circuitry have also been demonstrated.[111] Published research suggests that while considerable progress towards a minimal cell chassis has come a long way, there is much still to do before it is ready for wide-scale use.[112] There has also been significant progress in re-engineering standard research and industrial microbes, such as *E. coli* and *S. cerevisiae*, to make them more suitable for use as chassis.[113]

[95] http://www.nature.com/news/2010/100120/full/463288a.html
[96] http://www.jbioleng.org/content/3/1/4
[97] http://www.pnas.org/content/early/2012/04/20/1119407109.abstract
[98] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000167
[99] http://www.ploscompbiol.org/article/info:doi%2F10.1371%2Fjournal.pcbi.1000125
[100] http://www.biomedcentral.com/1752-0509/2/103
[101] http://www.ncbi.nlm.nih.gov/pubmed/18796402
[102] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2267732/
[103] http://www.jbioleng.org/content/4/1/14
[104] http://www.jbioleng.org/content/4/1/12
[105] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2671590/
[106] http://www.jbioleng.org/content/4/1/17
[107] http://www.jbioleng.org/content/3/1/19
[108] http://www.nature.com/news/2010/100722/full/news.2010.367.html
[109] http://www.nature.com/nbt/journal/v29/n1/full/nbt.1753.html
[110] http://www.nature.com/nchembio/journal/v5/n11/abs/nchembio.218.html
[111] http://www.jbioleng.org/content/4/1/3
[112] http://www.nature.com/msb/journal/v2/n1/full/msb4100090.html
[113] http://www.nsf.gov/news/news_summ.jsp?cntn_id=121639

18. The last few years have also seen the development of a range of different components that could be used with — or independently from — such chassis, including: rewired genetic switches;[114] functional molecules, such as re-engineered ribosomes; cell-free metabolic platforms for protein production;[115] non-natural synthetic proteins;[116] synthetic cell membranes;[117] as well as a self destruct mechanism to prevent engineered organisms surviving outside of laboratory settings.[118] A 2012 review of components included: regulatory cascades; epigenetic toggle switches; hysteretic circuits; molecular timing devices; synthetic eco-sensing systems; synthetic quorum-sensing systems; synthetic hormone systems; band-pass filets; as well as oscillators with tuneable frequency and amplitude.[119]

19. The same review noted that "a decade after the pioneering synthetic networks were reported, the first successful therapeutic applications in animal models of prominent human diseases are starting to emerge".[120] These studies include the "first synthetic closed-loop control gene network that manages homeostasis of a crucial disease metabolite in an animal model" and the "first optogenetic device that controls the production of a therapeutic protein in an animal disease model". It also examines other emerging biomedical applications, including for: understanding disease mechanisms, such as pathogen mechanisms and the immune system; disease prevention, such as vaccines and vector control; drug development, such as drug discovery, production and delivery; novel treatments for infections, such as breaking bacterial resistance and engineering pro-biotic bacteria to decrease pathogen virulence; cancer therapies, such as bacterial synthetic devices, viral synthetic devices and transformation sensors for cancer therapy; and other aspects, such as RNA controllers for cell proliferation, optogenetic devices in blood glucose homeostasis and prosthetic networks.

20. One challenge to the eventual wide-scale use of technology derived from synthetic biology will be the control of agents following release. Considerable work has already been undertaken to create kill switches designed to prevent undesirable spread.[121] Similar approaches are already yielding results in other fields.[122]

21. The safety and security implications of synthetic biology have been examined closely in parallel with scientific and technological developments.[123] Concerns have already been raised over military investment in synthetic biology.[124] Key reports published since 2006 include:

(a) *New Directions: The Ethics of Synthetic Biology and Emerging Technologies* by the Presidential Commission for the Study of Bioethical Issues in the United States; [125]

(b) *Synthetic Biology: the Technoscience and its Societal Consequences* by the SYNBIOSAFE project; [126]

[114] http://www.sciencedaily.com/releases/2010/01/100125173244.htm
[115] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2583083/
[116] http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0015364
[117] http://www.technologyreview.com/news/423381/making-cells-on-an-assembly-line/
[118] http://www.ncbi.nlm.nih.gov/pubmed/21645422
[119] http://www.nature.com/nrg/journal/v13/n1/abs/nrg3094.html
[120] http://www.nature.com/nrg/journal/v13/n1/abs/nrg3094.html
[121] http://www.pnas.org/content/early/2010/08/09/1009747107.abstract
[122] http://www.nejm.org/doi/full/10.1056/NEJMoa1106152
[123] http://www.livescience.com/10715-synthetic-biology-great-promise-potential-peril.html
[124] http://www.nature.com/news/bioengineers-debate-use-of-military-money-1.9409
[125] http://www.bioethics.gov/documents/synthetic-biology/PCSBI-Synthetic-Biology-Report-12.16.10.pdf

(c) Synthetic Biology: Social and Ethical Challenges by the Institute for Science and Society;[127]

(d) *Synthesis Report on Opportunities and Challenges in the Emerging Field of Synthetic Biology* by the Organization for Economic Cooperation and Development and the Royal Society; [128]

(e) *Risk Governance of Synthetic Biology* by the International Risk Governance Council; [129]

(f) *Synthetic Biology: Scope, Applications and Implications* by the Royal Academy of Engineering;[130]

(g) *What Rough Beast? Synthetic Biology, Uncertainty, and the Future of Biosecurity*, by academics at the Massachusetts Institute of Technology and Boston University; [131]

(h) *Security Implications of Synthetic Biology and Nanobiotechnology* by the United Nations Interregional Crime and Justice Institute (UNICRI);[132]

(i) *The Transnational Governance of Synthetic Biology: Scientific Uncertainty, Cross-borderness and the Art of Governance* by the London School of Economics and Political Science;[133] and

(j) *Synthetic Biology: Four Steps to Avoid a Synthetic Biology Disaster* by the Woodrow Wilson International Center for Scholars.[134]

In general, these reports recognise that synthetic biology "appears to have minimal security implications in the near term, create modest offensive advantages in the medium term, and strengthen defensive capabilities against natural and engineered biological threats and enable novel potential offensive uses in the long term".[135] Similar findings were echoed in the UNICRI review published in 2011.

IV. Gathering and manipulating biological information

22. Advances in bioinformatics and computational biology have greatly aided the gathering, processing and utility of biological data. Laboratories are becoming increasingly digitized.[136] This has helped extract information that was previously obscured and has made it easier and quicker to accomplish certain tasks. Increasingly the life sciences are referred to as information sciences. Digital tools and platforms not only support laboratory work but are increasingly able to replace it.

23. Descriptive languages developed over the last few years have included: a language for standardising and automating biology protocols: as well as a modelling language

[126] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2671589/
[127] http://www.bbsrc.ac.uk/web/FILES/Reviews/0806_synthetic_biology.pdf
[128] http://www.oecd.org/dataoecd/23/49/45144066.pdf
[129] http://www.irgc.org/IMG/pdf/IRGC_Concept_Note_Synthetic_Biology_191009_FINAL.pdf
[130] https://www.cbd.int/doc/emerging-issues/UK-submission-2011-013-Synthetic_biology-en.pdf
[131] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1452053
[132] http://igem.org/wiki/images/e/ec/UNICRI-synNanobio-final-2-public.pdf
[133] http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2011/4294977685.pdf
[134] http://www.nature.com/nature/journal/v483/n7387/full/483029a.html
[135] http://www.bioone.org/doi/abs/10.2990/28_2_2
[136] http://www.nature.com/news/going-paperless-the-digital-lab-1.9881

derived from one used in artificial intelligence that allows for better descriptions of biological processes. [137]

24. Advances in data mining have included: using multiple applications and datasets to reveal additional information about a system;[138] using Boolean logic to help identify genes; merging network theory and microarray data to reveal information about the co-expression of genes; [139] and tools for identifying interesting relationships between pairs of variables in large data sets.[140]

25. Capabilities in modelling and simulation have advanced significantly, including in: incorporating non-linear complexity, such as by adopting enzyme-centric approaches; as well as combining rule-based representations with agent-based simulation. [141]

26. It is now possible to recreate and in some cases make predictions from computational representations of: pathogenicity in fungi;[142] gene circuits, including filling in gaps that cannot be measures experimentally;[143] protein-protein interactions from amino acid sequence data and network structure;[144] [145] biochemical and diffusion reactions both in parts of cells and in whole cell contexts;[146] metabolic networks (including a model for the complete metabolic network of a pseudomonas)[147] with significant progress in simplifying networks,[148] modularizing them,[149] and better describing the dynamic nature of living cells;[150] cellular responses to external stimuli;[151] inter-cellular communication and cooperation with biomimetic microcapsules;[152] as well as whole-cell simulations for bacteria such as *E. coli* and *M. genitalium.*[153] [154]

27. Online tools made available over the past five years include: metabolic mapping software, for both whole metabolic networks and specific pathways;[155] platforms for comparative and functional genomics;[156] as well as the management and quality analysis of gene sequences.[157] Substantial investment has been made in developing new platforms designed to handle the volume of data produced by contemporary sequencing studies.[158]

28. Software suites are also available for use offline. Some of this software makes it easier to visualise complex biological information, including: genome sequence data:

[137] http://www.bioone.org/doi/abs/10.2990/28_2_2
[138] http://www.ncbi.nlm.nih.gov/pubmed/20231483
[139] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000117
[140] http://www.sciencemag.org/content/334/6062/1518
[141] http://www.biomedcentral.com/1752-0509/2/70
[142] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2387229/
[143] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2586632/
[144] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000118
[145] http://www.ncbi.nlm.nih.gov/pubmed/18277381
[146] http://www.biomedcentral.com/1752-0509/2/66
[147] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000210
[148] http://www.biomedcentral.com/1752-0509/2/86
[149] http://www.biomedcentral.com/1752-0509/2/78
[150] http://www.biomedcentral.com/1752-0509/2/84
[151] http://web.mit.edu/newsoffice/2011/vivo-systems-biology-0323.html
[152] http://www.pnas.org/content/107/28/12417.abstract
[153] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002010
[154] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000285
[155] http://nar.oxfordjournals.org/content/early/2011/05/28/nar.gkr433.full
[156] http://nar.oxfordjournals.org/content/early/2009/11/11/nar.gkp919.short
[157] http://www.biomedcentral.com/1471-2105/9/483/abstract
[158] http://www.genomeweb.com/informatics/nhgri-funds-new-sequencing-data-software-projects

sequence assembly data; plasmid maps; gene expression; comparative and functional genomic data; transcription; secondary structure of RNA;[159] and biochemical networks. [160]

29. Other software has been developed for gene sequence analysis, including for: basic analysis; structural analysis; comparative analysis; the identification of operons;[161] the identification of repeats; the identification of signalling-relevant motifs; the identification of protein coding genes: as well as links with metabolic function and disease.[162]

30. Protein analysis tools have been developed to: take advantage of power graph analysis; identify protein functional modules; as well as for sequence analysis.[163]

31. Other tools have been released to help: annotate genomes; model thermodynamics of reactions; analyse metabolomic data; and identify opportunities to repurpose drugs.[164] There have also been efforts to make use of machine learning capacity to: identify highly designable protein sequences;[165] and study and validate essential enzymes in a metabolic network.[166]

32. There has also been notable progress in moving from descriptive and analytical tools to design tools to assist in designing and conducting experiments.[167] Design tools released over the last few years include those for: gene design: sequence design; gene network design; plasmid design; PCR design; protein design; as well metabolic pathway design.[168]

33. Combining advances in bioinformatics with those in characterization as well as high-throughput approaches, and robotics is beginning to enable automated research approaches. Advanced modelling software has been used to take partially-characterised biological systems (such as those from yeast functional genomics or drug screening) and through the use of artificial intelligence develop theories as to what the missing components of the system might be (both in terms of intermediaries and processes). These computational models can then be tested through laboratory experimentation, where all the equipment is controlled by the same computer that developed the theories. Beyond restocking basic expendable laboratory resources, the experiments are conducted without human intervention. The same computer then assesses the outcomes of the experiments and feeds the data back into the model and uses it to improve its theories. This process is then repeated until the system is fully elucidated. The ability of robot scientists to characterise biological systems has been assessed through empirical study. The robot scientists were provided partial data from well characterised networks and asked to deduce the rest. Results from these studies indicated that the robot scientists are capable of characterising discrete biological systems.[169,170,171] Not only do robot scientists promise to take much of the drudgery out of basic research but they might also help to address the current bottlenecks in characterizing parts, identifying function and interpreting raw data.

159 http://gvi.seas.harvard.edu/paper/multeesum-tool-comparative-spatial-and-temporal-gene-expression-data
160 http://www.biomedcentral.com/1752-0509/2/104
161 http://genomebiology.com/2008/9/12/R179
162 http://www.biomedcentral.com/1752-0509/2/93
163 http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000108
164 http://www.biomedcentral.com/1471-2105/9/470
165 http://www.biomedcentral.com/1471-2105/9/487
166 http://www.biomedcentral.com/1471-2105/9/487
167 http://www.sciencemag.org/content/332/6031/816.abstract
168 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2238713/
169 http://www.sciencemag.org/content/324/5923/85.abstract
170 http://www.nature.com/nature/journal/v427/n6971/abs/nature02236.html
171 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1978088/

V. Converting biological information to digital data and back

34. If biology is becoming an information science then in part it is because of the ability to convert biological data into digital data and back again. Gene sequencing (reading the genetic code) enables us to move in one direction and gene synthesis (writing the genetic code) the other.[172] Capabilities to read and write genetic code are not new but capabilities in these areas have changed dramatically over the past five years.

35. Progress in sequencing has provided risks, benefits and challenges. It has added to the dual-use information previously available. For example, new pathogens, such as fungal plant pathogens, and the ricin-containing castor bean plant have been sequenced and the sequence data added to public databases. The same advances, however, help to strengthen public health capacity including molecular epidemiology, our understanding of pathogenesis, pathogen discovery and diagnosis, drug discovery, and vaccine development.[173] Recent events have demonstrated just how important this capacity can be. Advanced sequencing capacity enabled both the identification of an unknown agent responsible for a deadly disease outbreak in Germany in July 2011 and provided clues as to its origin and recent evolution. Increasing access to sequencing technology also raises the possibility of individuals having part (or all) of their genome sequenced and using the data to identify potential disease risks, which may in fact never be realised. Dealing with probabilities of disease is a complex task for highly trained medical professionals, allowing the general public access to such tools might well raise a series of conceptual, ethical and social challenges.[174]

36. Raw sequencing power has increased considerably over the past five years. Advances in technology continue to increase the throughput of automated gene sequencers. In December 2007, the Economist noted that a singe gene sequencer was capable of sequencing the human genome (about 3 billion nucleotides in length) in two months. A day's output from a first generation sequencer could be replicated, at the end of 2007, in less than 10 seconds. Second generation sequencers, such as 454 sequencing, provided "higher throughput, simplified all in vitro sample preparation and the miniaturization of sequencing chemistries, enabling massively parallel sequencing reactions to be carried out at a scale and cost not previously possible".[175] Over the intervening years two sets of sequencers Illumina (Illumina GA IIx SOLiD 3.0 and Illumina Hi-Seq 2000) used different massively parallel sequencing approaches to increase sequence output per instrument run by another order of magnitude.[176]

37. By early 2011, third generation *ion torrent* sequencing was possible. These US$50,000 machines "can read a bacterial genome in as little as two hours". The ion iorrent machine takes advantage of semiconductor manufacturing techniques and integrated circuits and "uses cheaper, natural nucleotides, and senses the hydrogen ions (protons) that are released as each nucleotide is incorporated onto the complementary DNA".[177] Current versions of ion torrent machines are not as accurate as some of their predecessors and might be "better suited to achieving fast results in smaller scale projects, such as sequencing bacterial genomes or characterizing diseases by reading certain gene regions across many

[172] http://www.rothamsted.ac.uk/ppi/pubs/kimhk/Beacham%20_et_al_2009_The_Biologist.pdf
[173] http://www.nejm.org/doi/full/10.1056/NEJMra1003071
[174] http://eon.businesswire.com/news/eon/20110727006628/en/Infectious-disease/pathogen-detection/e.-coli
[175] http://www.nature.com/nbt/journal/v26/n10/full/nbt1485.html
[176] http://www.nature.com/nature/journal/v470/n7333/full/nature09796.html
[177] http://www.nature.com/nature/journal/v475/n7356/full/nature10242.html

patients".[178] At least one version of the machine currently comes with an iPod dock. Next generation sequencers, such as those based on nanopore technology, are already under development and promise to cut costs and boost output even further.[179] The ion proton sequencer was released in January 2012. This, according to the manufacturers, can sequence an entire human genome in a day for $1000.[180]

38. A month later, rumours began to circulate of a new platform technology. Oxford Nanopore then announced the release of two machines the GridION and MinION. Both, according to the manufacturer, can read millions of bases per hour from samples with minimal preparations, including blood samples. The MinION is a disposable, memory-key sized unit which can be plugged into a computer for under $1000.[181]

39. Instrument output is not the only measure of progress in sequencing. The cost per base of sequencing has continued to fall. When the preliminary sequences of the human genome were released in 2000, they had cost millions of dollars. It was reported in the New Scientist in March 2008 that a commercial biotechnology company in California, USA had sequenced a human genome for $60,000, excluding labour. Over the past five years the cost per base has dropped by around four orders of magnitude. Advances in microfluidics look set to decrease the price even further. Equally, there are indications that the quality of sequence reads (in terms of lower error rates) have also gone up.[182] The current financial constraints and their impact on research funding could, however, reduce incentives that have driven recent advances.[183]

40. There are certainly rewards to be had for working on increased automation, accuracy and speed and decreased costs. In addition to the commercial applications, the X Prize Foundation,is now offering a $10 million prize for the first team to sequence 100 individual genomes with an accuracy of 99%, within 10 days. Each sequence is to contain at least 98% of the genome and cost $10,000 or less. [184]

41. This increased sequencing capacity has been used in a number of ways. It has enabled new types of projects to be attempted and as a result gathered different data sets,[185] including cataloguing sequences and their variation, assessing dynamic DNA and mixed genomes, investigating the epigenome and transcriptome, as well as combining different -omic approaches.

42. Health-related applications are increasingly common, for example, in diagnosing extremely rare genetic disorders,[186] working with hereditary conditions,[187] or infantile mitochondrial disease.[188] Over half of the genome sequences to date are part of disease specific projects.[189] For example, in 2001 the genome for the causative organism for plague was published throwing new light on the evolution of this pathogen.[190] Public funds are

178 http://www.nature.com/news/2011/110720/full/475278a.html
179 http://pubs.acs.org/doi/abs/10.1021/nl103873a
180 http://www.lifetechnologies.com/us/en/home/about-us/news-gallery/press-releases/2012/life-techologies-itroduces-the-bechtop-io-proto.html
181 http://www.nature.com/nbt/journal/v30/n4/full/nbt0412-295.html
182 http://www.technologyreview.com/news/419258/the-30-genome/
183 http://www.nature.com/news/2011/111101/full/479017a.html
184 http://www.technologyreview.com/news/419258/the-30-genome/
185 http://www.nature.com/nbt/journal/v26/n10/full/nbt1494.html
186 http://www.nature.com/news/2011/111005/full/478022a.html
187 http://www.technologyreview.com/review/412209/a-hole-in-the-genome/
188 http://stm.sciencemag.org/content/4/118/118ra10.abstract
189 http://www.pnas.org/content/108/4/1513.full
190 http://www.nature.com/nature/journal/v478/n7370/full/nature10549.html

being invested to develop medical applications based on advanced sequencing capacity.[191] Companies and service providers have already begun to work on tools and platforms.[192,193]

43. Advanced sequencing capacity can be found on every continent and, in line with broader trends in biotechnology, increasingly in developing countries. An interactive map created by the Bacterial Pathegonomics research group at the University of Birmingham in the United Kingdom illustrates the global spread.[194]

44. Despite the distribution of sequencers, there is less geographical balance in the genes being sequenced. There has been an exponential growth in the number of human genomes that have been sequenced. Only two had been sequenced at the Sixth Review Conference in 2006. By the end of 2011, it was estimated that over 30,000 human genomes had been sequenced. The majority of these, however, are from Caucasian or Asian individuals; very few African and South American genomes have been complete.[195] Similar disparities exist in medical genomics and there have been calls to expand the sequencing of other ethnic groups.

45. There has also been progress in ability to understand and use sequence data. Genome mining techniques have started to identify useful compounds encoded within sequence data. [196]Genome wide analysis and association studies have: improved linkages between sequence data and metabolomics data; linked genetic variations at specific loci with particular diseases;[197] led to personal genome scans which can provide risk indicators for specific diseases;[198] and provided insights into mutation rates. Deep sequencing has also made steady headway in helping to determine gene function.[199]

46. Trends in synthesis capacity mirror those for sequencing. There have been technical improvements in the ability to produce longer strands of genetic material. New assembly techniques make is easier and faster to combine short fragments into long sequences.[200] These techniques were used in 2010 to build a piece of DNA with over one million base pairs. The cost of having gene length fragments commercially synthesized also continues to fall (even faster than the costs of synthesizing smaller oligonucleotide sequences).[201] Quality seems to be increasing, with both recursive and re-sequencing approaches providing for more effective error correction.[202] For example, in February 2012, Integrated DNA Technologies introduced a new service, which it claims will deliver double-stranded, sequence verified, genomic blocks up to 500bp within 3-4 working days with a 33% decrease in costs over similar services in the past.[203] Days later the company announced a new partnership with Synthetic Genomics to use this platform to offer commercial production of custom, synthetic, double-stranded genomic fragments up to 5000 base pairs.[204]

---

191 http://www.nature.com/news/funds-dedicated-to-personalized-genetics-1.9565

192 http://www.genomeweb.com/sequencing/life-tech-opgen-combine-technologies-outbreak-surveillance

193 http://www.guardian.co.uk/science/2011/dec/28/mayo-clinic-genomes-personalised-care

194 http://pathogenomics.bham.ac.uk/hts/

195 http://www.nature.com/nature/journal/v456/n7218/full/456049a.html

196 http://www.microbeworld.org/index.php?option=com_jlibrary&view=article&id=4343

197 http://www.nature.com/nature/journal/v477/n7362/full/nature10354.html

198 http://www.nature.com/nature/journal/v456/n7223/full/nature07631.html

199 http://www.ncbi.nlm.nih.gov/pubmed/21623355

200 http://www.nature.com/nmeth/journal/v6/n5/full/nmeth.1318.html

201 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2424292/

202 http://www.nature.com/nmeth/journal/v8/n2/full/nmeth0211-114.html

203 http://eu.idtdna.com/pages/mobile/news/2012/01/31/integrated-dna-technologies-introduces-gblockstm-gene-fragments

204 http://manufacturing.pharmaceutical-business-review.com/news/sgi-idt-partner-to-manufacture-

47. The projects being attempted with synthesis technologies are also becoming more sophisticated. At the time of the last review conference, cutting edge application was taking place in viral settings, May 2010 saw the chemical synthesis of a functional genome capable of controlling a bacterial cell,[205] and by November 2010 similar approaches were being used in animal models (although to chemically synthesize the genome of mitochondria, not the mouse in which it is found).[206] By September 2011, this had moved again to synthesis of part of the chromosome of a eukaryote. [207]

VI. Generic enabling technologies

48. Underpinning many of the advances discussed throughout this paper are a range of technologies that make it easier, cheaper, faster or more reliable to do many of the basic procedures and practices involved in expanding the limits of understanding and creating new applications. Other advances have allowed scientists to do things that were previously unattainable.[208] Significant enabling technologies developed over the past five years included:

(a) A simpler, cheaper and more reliable way of forming carbon-hydrogen bonds important in biochemical synthesis;[209]

(b) Gene profiling and agent identification using quantitative PCR;[210]

(c) Faster and more accurate ways of determining the three dimensional structure of biological macromolecules using new synchrotron light sources;[211]

(d) Tools to study the binding and unbinding of individual strands of DNA through a combination of flourescent microspopy and optical traps;[212]

(e) An high-throughput tool for in vivo analysis of bioactive small molecules important for modulating protein function and important leads for drug discovery;[213]

(f) New ways to create diverse small molecule drug candidate libraries enabling high-throughput drug discovery;[214]

(g) Real-time, multi-parameter analysis of single immune cells using single cell mass cytometry (tools used to make measurement of impurities in superconductors);[215]

(h) Better imaging tools, including digital holographic microscopes,[216] three-dimensional isotrophic imaging of living cells using Bessel beam place illumination,[217] as well as sub-diffraction-limit imaging by stochastic optical reconstruction microscopy

synthetic-gene-products-030212

205 http://www.sciencemag.org/content/329/5987/52.abstract

206 http://www.nature.com/nmeth/journal/v7/n11/full/nmeth.1515.html

207 http://www.ncbi.nlm.nih.gov/pubmed/21918511

208 http://www.nap.edu/catalog.php?record_id=12601

209 http://www.scripps.edu/news/press/2009/120309.html

210 http://www.nature.com/nmeth/journal/v8/n3/full/nmeth0311-207.html

211 http://connection.ebscohost.com/c/articles/59207776/illuminating-science-how-synchrotrons-are-revolutionising-structural-biology

212 http://news.illinois.edu/news/11/0302DNA_TKHa_YannChemla.html

213 http://www.ncbi.nlm.nih.gov/pubmed/18622389

214 http://www.nature.com/nature/journal/v457/n7226/full/457153a.html

215 http://www.sciencemag.org/content/332/6030/687.abstract

216 http://www.nap.edu/catalog.php?record_id=12821

217 http://www.ncbi.nlm.nih.gov/pubmed/21378978

(STORM), which enables the simultaneous imaging of multiple molecules in living cells and has been used to examine the changes in concentration of proteins in the membranes of immune cells when they encounter toxins;[218]

(i) Improvements in temporal analysis of gene expression using short-time series microarrays which enable expression to be tracked more accurately over time, perhaps as a system is perturbed; [219]

(j) A way to specifically target endogenous gene sequences to introduce mutations, tags or new sequences via optimized transcription-activator-like effector (TALEs);[220]

(k) Tools for single cell analysis, including its genome, transcriptome, metabolome, and peptidome; [221]

(l) The use of quantum dots to tag and track individual viruses;[222]

(m) A much faster and simplified way of compiling short sections of genetic material together to make longer strands, via the Gibson assembly technique;[223]

(n) Better optimized protein production in *E.coli* through continuous directed evolution of gene encoded molecules via phage-assisted continuous evolution (PACE);[224]

(o) Genome editing tools for small-scale genome engineering by the programming and evolution of cells by simultaneously targeting many locations on their chromosome via multiplex automated genome engineering (MAGE)[225] and MAGE codon modifications to provide for large-scale genome via hierarchical conjugative assembly genome engineering (CAGE);[226] [227]

(p) Inserting genetic material into cells, by either using a gene gun (which was created prior to the last review conference but has been improved considerably since) or via a non-viral plasmid;[228]

(q) More sophisticated microfluidic applications, such as the addition of optical pumps or better system integration, which improves the utility of a lab-on-a-chip;[229]

(r) Cell free systems designed to produce encoded proteins from synthesised DNA via nucleic acid programmable protein arrays (NAPPA);[230]

(s) A way to control cell function using light (which provides targeted, fast control of precisely defined events in biological systems) through optogenetics;[231]

[218] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2700296/
[219] http://www.biomedcentral.com/1752-0509/2/58
[220] http://www.nature.com/nmeth/journal/v8/n3/full/nmeth0311-197.html
[221] http://www.nature.com/nmeth/journal/v8/n4s/full/nmeth0411-S1.html
[222] http://www.newscientist.com/article/dn14675-viral-manoeuvres-revealed-by-surveillance-system.html
[223] http://www.nature.com/nmeth/journal/v6/n5/full/nmeth.1318.html
[224] http://www.nature.com/nbt/journal/v29/n6/full/nbt.1884.html
[225] http://nextbigfuture.com/2010/08/george-churchs-multiplex-automated.html
[226] http://phys.org/news/2011-07-genome.html
[227] http://www.sciencemag.org/content/333/6040/348.abstract
[228] http://discover-decouvrir.cisti-icist.nrc-cnrc.gc.ca/eng/article/?id=17719349
[229] http://www.ncbi.nlm.nih.gov/pubmed/21612614
[230] http://nextbigfuture.com/2008/02/any-protein-can-be-made-from.html
[231] http://www.nature.com/nmeth/journal/v8/n1/abs/nmeth.f.325.html

(t) Approaches for tissue engineering and assembling three dimensional biological structures and using standardised blocks or through printing; [232]

(u) Automated research suites designed to enable high-throughput screening campaigns, including those intended for use under BSL-2 conditions;[233]

(v) Increasingly comprehensive sets of normal data stored in biobanks, including genetic information and blood samples as well as medical and family histories;[234]

(w) A new way to trap and manipulate micro-scale objects using mobile micro-vortices;[235]

(x) A protocol for using multi-isotope imaging mass spectrometry (MIMS) in living cells at the sub-micrometer range;[236]

(y) A new method for assessing the "drug-likeness" of compounds;[237]

(z) High-throughput screening tools to screen libraries of compounds for biological activity based upon improvements in microfluidics.[238]

---

[232] http://web.mit.edu/newsoffice/2010/tissue-legos-0513.html
[233] http://www.highresbio.com/pdf/HighRes_Bio_in_NatureMethods0908_843.pdf
[234] http://www.nature.com/nbt/journal/v29/n6/full/nbt.1884.html
[235] http://pubs.acs.org/doi/abs/10.1021/nl2032487
[236] http://www.nature.com/nature/journal/v481/n7382/full/481454a.html
[237] http://www.nature.com/nature/journal/v481/n7382/full/481455a.html
[238] http://www.nature.com/nature/journal/v483/n7387/full/483043a.html