International Standard

ISO 20688-2

First edition
2024-03

Biotechnology — Nucleic acid synthesis —

Part 2: Requirements for the production and quality control of synthesized gene fragments, genes, and genomes

Reference number
ISO 20688-2:2024(en)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Requirements for quality management **3**
4.1 General requirements 3
4.2 Control of documents 4
4.3 Quality management system 5
4.4 Biorisk management and safety control 5
5 Requirements for resource management **5**
5.1 Facilities and environmental condition 5
5.2 Equipment and instruments 5
5.3 Raw materials 6
5.4 Personnel 6
6 Biosafety and biosecurity requirements **6**
6.1 General 6
6.2 DNA sequence screening mechanism 6
7 Requirements for quality control in production **7**
7.1 General 7
7.2 Quality control in synthetic gene fragments production 7
7.2.1 General 7
7.2.2 Sequence design 7
7.2.3 Assembly 8
7.2.4 Purification 8
7.2.5 Product preservation 8
7.3 Quality control in synthetic gene production 8
7.3.1 General 8
7.3.2 Colony screening 8
7.3.3 DNA preparation from the host cell 8
7.3.4 Sequence verification 9
7.3.5 Product preservation 9
7.4 Quality control in synthetic genome production 9
7.4.1 General 9
7.4.2 Assembly 9
7.4.3 Sequence verification 9
8 Requirements for product quality **10**
8.1 Synthetic gene fragments 10
8.1.1 General 10
8.1.2 Yield 10
8.1.3 Purity 10
8.1.4 Size 10
8.1.5 Gene cloning accuracy 10
8.2 Synthetic genes 10
8.2.1 General 10
8.2.2 Yield 10
8.2.3 Purity 11
8.2.4 Sequence 11
8.2.5 Integrity 11
8.2.6 Residual impurities analysis 11
8.2.7 Supercoiled plasmid 12
8.2.8 Bioburden 12

8.2.9 Specific quality indicators for synthetic DNA libraries ... 12
8.3 Synthetic genome ... 12
8.3.1 General ... 12
8.3.2 Sequence ... 12
9 Delivered/synthesized material specifications ... 13
9.1 Main information ... 13
9.2 Other information ... 13
Annex A (informative) **Example equipment and device list and their control requirements ... 14**
Annex B (informative) **Exemplary quality measurement methods ... 16**
Annex C (informative) **Electrophoretogram ... 18**
Annex D (informative) **Sanger sequencing ... 22**
Annex E (informative) **Massively parallel sequencing ... 23**
Annex F (informative) **Additional quality control options for synthetic gene fragments ... 26**
Annex G (informative) **Example ranking of the risk posed by DNA products ... 28**
Bibliography ... 29

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 276, *Biotechnology*.

A list of all parts in the ISO 20688 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Gene fragment, gene and genome synthesis refer to producing synthetic double-stranded DNA in the form of non-clonal fragments (that can be linear) and clonal genes in plasmids (that would be circular) by using appropriate biochemical methods.

Synthesized gene fragments, genes and genomes are important biotechnological products and are widely used in biotechnology, e.g. protein engineering, metabolic engineering, antibody and vaccine development, environmental bioremediation and natural product discovery.

The production and quality control of the synthesized gene fragment, gene and genome products are essential for ensuring the quality and their downstream applications in biotechnology. This document provides requirements for the production and quality control of synthetic gene fragment, gene and genome products, including biosecurity, purity, yield, size, gene cloning accuracy, integrity, sequences, residual impurities and other quality indicators. This document provides a uniform general guideline for the quality control of gene fragment, gene and genome synthesis. It is intended to help to improve and ensure the quality of products and fair trade based on a unified standard.

This document is intended to be used by synthetic DNA producers during the manufacturing process for quality control to improve the quality of their products, by academic laboratories to evaluate the quality of DNA synthesized in their facilities, and by end users to verify the quality of synthesized gene fragments, genes and genomes provided by manufacturers as required.

In this document, the following verbal forms are used:

— “shall” indicates a requirement;

— “should” indicates a recommendation;

— “may” indicates a permission;

— “can” indicates a possibility or a capability.

Biotechnology — Nucleic acid synthesis —

Part 2: Requirements for the production and quality control of synthesized gene fragments, genes, and genomes

1 Scope

This document specifies the requirements for the production and quality control of synthesized double-stranded DNA. It describes requirements for quality management, resource management, biosafety and biosecurity, quality control in production, product quality, and delivered product specifications for synthesized gene fragments, genes and genomes.

This document is applicable to synthetic gene fragments, genes and genomes with a length below 10 Mbp (base pairs) in the forms of non-clonal fragments (linear) and clonal genes in plasmids (circular).

This document does not provide specific requirements for materials used solely for diagnostic purposes.

When the synthesized nucleic acids are procured and used for diagnostic purposes, the user can take ISO 15189, ISO 13485 and other related clinical standards into account.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
biosafety
practices and controls that reduce the risk of unintentional exposure or release of biological materials

Note 1 to entry: Biological materials refer to any material comprised of, containing, or that may contain biological agents and/or their harmful products, such as toxins and allergens (see ISO 35001:2019, 3.14).

Note 2 to entry: Biological agents refer to any microbiological entity, cellular or non-cellular, naturally occurring or engineered, capable of replication or of transferring genetic material that may be able to provoke infection, allergy, toxicity or other adverse effects in humans, animals, or plants (see ISO 35001:2019, 3.13).

[SOURCE: ISO 35001:2019, 3.22, modified — Notes to entry were added.]

3.2
biosecurity
practices and controls that reduce the risk of loss, theft, misuse, diversion of, or intentional unauthorized release of biological materials

[SOURCE: ISO 35001:2019, 3.23, modified — Notes to entry were deleted.]

3.3
colony polymerase chain reaction
colony PCR
PCR method used to screen for plasmids containing a desired insert directly from microbial colonies without plasmid extraction and purification steps

3.4
DNA assembly
joining oligonucleotides or smaller gene fragments via regions of complementarity to form a longer double-stranded DNA fragment step by step *in vitro* or *in vivo*

3.5
DNA sequencing
determining the order of nucleotide bases (adenine, guanine, cytosine and thymine) in a molecule of DNA

Note 1 to entry: Sequence is generally described from the 5′ end.

[SOURCE: ISO 17822:2020, 3.19]

3.6
gene cloning
process of introducing a particular gene or DNA sequence using genetic engineering techniques into a host cell and replicating it by asexual reproduction into many identical copies of the gene

3.7
massively parallel sequencing
MPS
sequencing technique based on the determination of incremental template-based polymerization of many independent DNA molecules simultaneously

Note 1 to entry: Massively parallel sequencing technology can provide millions/billions of short reads per run or long reads based on amplification.

[SOURCE: ISO 20397-2:2021, 3.30, modified — Note to entry was edited by adding "or long reads based on amplification.]

3.8
plasmid vector
extrachromosomal DNA molecule in cells physically separated from the chromosome and capable of autonomous replication that can be used as vehicle to carry new genes into cells

[SOURCE: ISO 16577:2022, 3.4.37, modified — “vector” added to the term, and “that can be used as vehicle to carry new genes into cells” added to the definition (from ISO 16577:2022, 3.4.58). Notes to entry have been deleted.]

3.9
quality score
Q score
measure of the sequencing quality of a given nucleotide base

Note 1 to entry: Q is defined by the following formula:

$$Q=-10\log_{10}(p)$$

where p is the estimated probability of the base call being wrong.

Note 2 to entry: A quality score of 20 represents an error rate of 1 in 100, with a corresponding call accuracy of 99 %.

Note 3 to entry: Higher quality scores indicate a smaller probability of error. Lower quality scores can result in a significant portion of the reads being unusable. Low quality scores can also indicate false-positive variant calls, resulting in inaccurate conclusions.

[SOURCE: ISO 20397-2:2021, 3.32]

3.10
sequence alignment
arrangement of nucleic acid sequences according to regions of similarity

Note 1 to entry: Sequence alignment may not require a reference genome/reference targeted nucleic acid region and its aim might not produce an assembly.

[SOURCE: ISO 20397-2:2021, 3.20]

3.11
sequence of concern
SOC
sequence of 50 bp or greater that either encode for biological functions or directly endow or enhance toxicity or pathogenicity

3.12
synthetic DNA library
double-stranded DNA fragments synthesized with targeted genetic diversity that have been inserted into a specific cloning vector(s)

Note 1 to entry: Genetic diversity refers to the number of unique sequences in a DNA library. Diverse libraries can permit high-throughput evaluation of genetic designs or functional variants.

3.13
synthetic gene
synthetic, cloned, double-stranded DNA fragment containing necessary biological parts

Note 1 to entry: Linear plasmid by restriction enzyme digestion is a kind of delivered product form of synthetic gene.

3.14
synthetic gene fragment
synthetic, non-cloned, double-stranded linear DNA fragments, assembled from synthetic oligo nucleotides

3.15
synthetic genome
synthetically-built genome containing all necessary genetic information for a living organism, produced by the assembly of oligonucleotides or smaller gene fragments *in vitro* or *in vivo*

4 Requirements for quality management

4.1 General requirements

The producer, as an entity synthesizing double-stranded DNA and distributing double-stranded DNA to one or more customer(s), shall establish and implement a system in which the following processes are described and documented:

a) order receiving process;

b) biosafety and biosecurity risk assessment process;

c) gene synthesis process;

d) final product quality control process.

A quality policy and quality objectives shall be determined in the order receiving process. The quality requirements are different depending on the synthetic production process, the form of the final synthetic double-stranded DNA, the quality control method and its end application. Necessary actions should be taken in the synthetic processes in order to achieve the planned results and quality by analysis of the characteristics of synthetic nucleic acids that are produced.

4.2 Control of documents

The producer of synthetic double-stranded DNA should have a procedure ensuring the control of documented information including the following points:

a) customer information:
   1) point-of-contact name;
   2) organization;
   3) address;
   4) phone number;
   5) email;

b) order sequence information:
   1) nucleotide sequences ordered;
   2) vector used;

c) sequence screening protocol;

d) sequence screening report;

e) standard operation procedure of synthesis;

f) quality control method;

g) product form;

h) data and report;

i) shipment information:
   1) date placed and shipped;
   2) shipping address;
   3) receiver name;
   4) transport storage conditions, etc.

The producer shall ensure that the unintended use of any obsolete document is prevented. The producer shall ensure the integrity and security of synthetic gene order and customer information and prevent unauthorized access to these data.

When the documented information including records is retained in electronic media, the producer shall ensure the control of those electronic media. Adequate cybersecurity measures shall be implemented to protect the intellectual property and identity of customers.

4.3 Quality management system

The producer can adopt and establish a quality management system to document necessary procedures, ensure control of production processes, and regularly monitor and document the production and quality control of synthetic double-stranded DNA.

4.4 Biorisk management and safety control

The producer can establish a biorisk management system (e.g. based on ISO 35001, the World Health Organization's (WHO's) *Laboratory Biosafety Manual*[10] and the WHO's *Global guidance framework for the responsible use of the life sciences*[11]) to effectively identify, assess, control, and evaluate the biosafety and biosecurity risks inherent in its activities.

The producer can establish an occupational health and safety management system (e.g. ISO 45001) in order to reduce or eliminate possible risks associated with performing double-stranded DNA synthesis and quality control as specified by this document.

The sequence of the synthetic double-stranded DNA should be screened against a list of pathogens and toxins. The biosafety and biosecurity risk level of the synthetic gene should be assessed according to the appropriate reference standard and documents of biosafety and biosecurity. An example of ranking risk levels can be referred to in Annex G.

The producer should have a procedure to ensure the legitimacy of customers, principal users and end users of synthetic genes containing sequences of concern (SOCs). Providers and third-party vendors of synthetic genes should:

a) know to whom they are distributing a product;

b) know if the product that they are synthesizing and/or distributing contains, in part or in whole, SOCs;

c) notify customers and end-users when their order contains SOCs.

5 Requirements for resource management

5.1 Facilities and environmental condition

Facilities, including sources of energy, lighting and environmental conditions (temperature, humidity, cleanness and atmospheric pressure) shall be functional and reliable for double-stranded DNA synthesis and quality control. Facilities and environmental conditions shall not adversely affect the synthesis and quality control of synthetic double-stranded DNA. Influences that can adversely affect the product quality can include, but are not limited to, other nucleic acid contamination, microbial contamination, dust, electromagnetic disturbances, radiation, humidity, inconsistent electrical supply, temperature and vibration.

Synthetic gene fragments, genes and genomes shall not be contaminated by other nucleic acids from the manufacturing environment and shall not be released into the exterior environment without proper treatment.

The producer shall monitor, control and record environmental conditions in accordance with relevant specifications, methods or procedures.

5.2 Equipment and instruments

Equipment and instruments used in the production and quality control of synthetic gene fragments, genes and genomes shall be properly controlled, maintained and calibrated.

The records of the control, maintenance and calibration shall be retained according to documented record retention policies.

The equipment and instruments shall be operated by suitably trained and qualified personnel.

The equipment for the production and quality control may include automated oligonucleotide synthesizers based on, but not limited to, column-based, microfluidics- and microarray-based devices. Additional necessary equipment may include polymerase chain reaction thermocyclers, gel electrophoresis apparatus, microchip capillary electrophoresis analysers, ultraviolet spectrophotometers, fluorescence spectrophotometers, DNA sequencers, centrifuges, incubators, refrigerators, freezers, pure water production systems, pH meters, weighing devices, pipettes, automated pipetting systems, dryers, constant temperature incubators, constant temperature shakers, etc.

An example of equipment and device list and their control requirements for the production and quality control of synthetic gene fragments, genes and genomes is given in Annex A.

5.3 Raw materials

Raw materials include synthetic substrates, auxiliary materials (such as reaction container, pipette tips, etc.), auxiliary reagents (such as oligonucleotides, enzymes, vector, culture medium, buffer, etc.) and pure water. Their quality will affect the quality of synthetic gene fragments, genes and genomes and the consistency and stability of manufacturing processes.

The producer shall control raw materials used in the production and quality control of synthetic gene fragments, genes and genomes.

The producer shall evaluate raw material suppliers to minimize the influences of the provided raw material on the synthesis requirements, such as purity of synthetic substrates, activity of enzymes, and so on.

The producer shall establish procedures for purchasing raw materials and evaluating suppliers against predetermined criteria. The producer should establish a process of reagent lot qualification to ensure that each reagent lot meets the requirements of the producer before use in manufacturing.

Only chemicals and materials of molecular biology grade shall be used.

5.4 Personnel

The producer should develop training programmes designed to ensure the competency required for manufacturing roles. The producer should provide training for all personnel according to the responsibilities assumed. The training programmes can include, but are not limited to, knowledge of chemistry, molecular biology and cellular biology.

6 Biosafety and biosecurity requirements

6.1 General

A risk- and evidence-based approach to biosafety shall be established and applied to ensure that laboratory facilities, safety equipment and work practices are locally relevant, proportionate and sustainable while maintaining appropriate control of biosafety.

Laboratory biosecurity measures should be taken based on a comprehensive programme of accountability for biological materials used in the production and quality control of synthetic gene fragments, genes and genomes.

In order to prevent the intentional or inadvertent misuse of DNA synthesis technologies and products, producers should make use of a DNA sequence screening mechanism.

For biosafety considerations, ISO 35001 and the WHO's *Laboratory Biosafety Manual*[10] can be used.

6.2 DNA sequence screening mechanism

All DNA producers should use a sequence screening mechanism to evaluate ordered sequences. This screening mechanism may be constructed in-house by producers or acquired from a third party. Screening systems can rely on an internationally recognized database of sequences of pathogen and toxin DNA and

algorithms to screen ordered DNA sequences against that set of sequences. Screening should be conducted for sequences longer than 50 bp or in accordance with regional guidelines. If the screening system returns a hit for an ordered DNA sequence, the DNA producer shall choose whether to conduct follow-up screening or to reject the order. Where follow-up screening does not resolve concerns about an order, the producer may choose to refuse the order or to report the order to authorities according to the particular case. For DNA producers that choose not to synthesize pathogen or toxin DNA, the synthesis should not proceed [13].

If a DNA producer chooses to synthesize pathogen or toxin DNA (i.e. sequences that are hits according to their screening mechanism), the producer shall follow legitimate use guidelines. Evidence for legitimate use may include institutional affiliation, evidence of a legitimate research programme, customer publication history or marketed products (e.g. detection and test).

If a DNA producer chooses to synthesize pathogen or toxin DNA, the producer shall establish the corresponding capacity and facility for maintaining an appropriate control of biosafety and biosecurity.

When a customer orders DNA sequences from a regulated pathogen or toxin, the producer shall obtain a written description of the intended use for the synthetic DNA from the customer.

Whenever possible, the producer shall verify that the information obtained, including the intended use, is consistent with the customer's activities. The result of the evaluation shall be documented.

It is recommended that producers document and retain for at least eight years the following information for orders about DNA sequences from a regulated pathogen or toxin:

a) customer information (point-of-contact name, organization, address, email, and phone number);

b) order sequence information (nucleotide sequences ordered, vector used);

c) order information (date placed and shipped, shipping address, receiver name).

d) intended use information (description from the customer, evaluation result)

7 Requirements for quality control in production

7.1 General

The producer shall establish a quality control system for the manufacturing of synthetic gene fragments, synthetic genes and synthetic genomes to ensure reliability and reproducibility. The quality control system shall establish a quality policy, quality objectives and necessary procedures to ensure the execution of manufacturing and quality control based on the established procedures.

7.2 Quality control in synthetic gene fragments production

7.2.1 General

Synthetic gene fragments are usually synthesized by assembling synthetic oligo nucleotides into double-stranded linear DNA fragments with appropriate assembly strategy and purified further when necessary.

EXAMPLE Assembly of full-length gene constructs for protein expression and purification, gRNA expression cassettes for CRISPR/Cas9-based gene editing, donor constructs for gene editing experiments and template for *in vitro* transcription.

7.2.2 Sequence design

Ordered sequences shall be evaluated to confirm whether the sequence can be synthesized correctly by the producer. Criteria to evaluate can include GC content, secondary structure or sequence repetitions. If sequences contain motifs that create manufacturing risk, codon optimization may be used to reduce manufacturing risk and/or to optimize protein expression in a desired host.

7.2.3 Assembly

An appropriate assembly strategy (including, but not limited to, polymerase chain reaction assembly (PCA), ligase chain reaction assembly (LCA), or recombination-based *in vitro*, *in vivo* assembly, etc.) can be applied to gene fragments synthesis to meet requirements of the final ordered sequence.

7.2.4 Purification

Assembly products can be purified by an appropriate method including PCR product purification kits, gel extraction purification kits, etc. Kits should be selected with the final purity requirements for the ordered synthetic DNA. Methods shall be validated when developed in-house or modified by the producer. The performance of purification methods shall be verified before use.

7.2.5 Product preservation

Synthetic gene fragments are shipped in formats including lyophilized powder or suspended in liquid buffer.

For lyophilized powder or product on certain medium such as filter paper, the product should be stored at 4 °C or below and may be temporarily stored at room temperature.

Gene fragments suspended in buffer should be stored below –20 °C. Multiple freezing and thawing cycles should be avoided.

7.3 Quality control in synthetic gene production

7.3.1 General

In the field of pharmaceutical development (e.g. viral genomes), metabolic engineering, pathway engineering, humanized antibodies production, or synthetic biology, longer synthetic genes in purified plasmid or glycerol stock of clonal cells containing the synthesized gene in a plasmid are needed.

Synthesized oligonucleotides serve as building blocks and are assembled into gene-length sequences. Two or more gene-length sequences can be further assembled into larger constructs. Final constructs may be inserted into plasmids for cloning. Factors which can affect assembly and plasmid replication, such as gene toxicity, regulatory elements and background expression, should be considered. If possible, all constructs, or at least the final plasmid construct, should be sequence-verified. Additional validation approaches such as enzymatic digestion or specialized DNA sequencing techniques shall be included for synthetic constructs with repetitive structures or stable secondary structures which are incompatible with common sequencing validation technologies. Clones containing perfect plasmid DNA (matching the ordered sequence with 100 % accuracy) may be shipped to customers as glycerol stocks or fresh stocks or may be further extracted to isolate the plasmid DNA. Plasmid DNA suspended in buffer or lyophilized form should then be delivered to the customer. Linearization by enzyme digestion may also be performed if customers prefer linearized, clonal DNA.

7.3.2 Colony screening

To screen colonies quickly, one or more methods may be used including colony PCR and antibiotic resistance screening to obtain properly transformed colonies.

7.3.3 DNA preparation from the host cell

When the synthesized genes are replicated in a specific host cell, the producer should select appropriate methods to extract and purify synthesized DNA (options include mini-column purification, magnetic bead-based purification, ethanol precipitation, phenol-chloroform extraction, etc.) from the host cell.

7.3.4 Sequence verification

The sequence of synthetic genes should be verified. Sequencing results shall be documented. The technology, such as Sanger sequencing, MPS or other equivalent DNA sequencing technology, should be selected according to gene features, synthesis throughput and cost considerations.

7.3.5 Product preservation

Synthetic genes may be shipped lyophilized, suspended in buffer solution, as glycerol stock or as fresh stock.

For lyophilized powder or product on certain medium such as filter paper, the product should be stored at 4 °C and may be momentarily stored at room temperature.

For DNA suspended in buffer solution, the product should be stored below –20 °C. Multiple freezing and thawing cycles should be avoided.

Synthetic genes in glycerol stocks of transformed cells such as *Escherichia coli.* should be stored below –20 °C and preferably at –80 °C.

Fresh stocks of transformed cells should be stored at 4 °C.

7.4 Quality control in synthetic genome production

7.4.1 General

Synthetic genome refers to re-designing, copying and the synthesis of an entire genome or all DNA sequences from a specified target organism. Sequence consistency of synthetic genome products is critical and shall be sequence-verified.

The entire process of constructing a synthetic genome shall be carried out in the following steps:

— analysing the entire genomic DNA sequence of the target organism;

— synthesizing individual fragments of an entire genome;

— assembling the genomic fragments;

— transplanting the assembled synthetic genome into the target organism;

— booting up the cell to express the genomic content.

Synthetic genome technology is used to create some of economically important microbes such as biofuel- or alcohol-producing microbes, which are used for decontaminating toxic waste or tracking down tumour cells.

7.4.2 Assembly

An appropriate assembly strategy should be selected according to the length of the genome. For relatively short sequences, polymerase-based, ligase-based or recombination-based *in vitro* and *in vivo* assembly methods can be used. Efficient assembly requires a high-fidelity enzyme (e.g. DNA polymerase or ligase). For larger DNA assemblies, seamless assembly methods that do not leave scars at the assembly junctions can be used.

7.4.3 Sequence verification

Due to the inherent potential for errors in each step of gene synthesis, including oligo synthesis and assembly, internal insertions and deletions as well as premature termination are inevitable in synthetic DNA sequences. Sequences should be verified before use in genome-scale assembly.

Sequences harbouring mutations shall be identified and removed. For unavoidable mutations, the mutation information validated by DNA sequencing should be provided to the customer, confirming with the customer whether the observed mutations affect downstream use or still meet the requirements of the customer.

Synthesized sequences are cloned into a plasmid vector in *Escherichia coli* or yeast and then sequenced by Sanger sequencing or MPS method.

8 Requirements for product quality

8.1 Synthetic gene fragments

8.1.1 General

The yield, purity and size of the delivered synthetic linear gene fragments products are important quality indicators.

8.1.2 Yield

The yield of synthesized linear gene fragments can be determined by measuring the concentration of the re-suspended products with different approaches, such as ultraviolet spectrophotometry via determination of optical density at 260 nm (OD_{260}) or via fluorescence intensity quantitation using DNA-binding fluorescent dyes. Quantitative PCR (qPCR) can be used as an additional validation method if the customer disputes the results. Refer to B.2.

For the ultraviolet absorption method, attention should be paid to the possible interference of unincorporated nucleotides and single-stranded nucleic acids.

8.1.3 Purity

Impurities in synthetic linear gene fragments can include proteins coming from the polymerase or ligase used for assembly. The ratio of optical density at 260 nm versus 280 nm can be used to assess protein contamination. The value of optical density (OD) in the range of 1,8 to 2,0 indicates little protein contamination in the product. Refer to B.1.

8.1.4 Size

The size of linear gene fragment products should be verified by agarose gel electrophoresis or capillary electrophoresis. The molecular mass of the measured bands in the electrophoresis shall be consistent with the ordered sequence length.

8.1.5 Gene cloning accuracy

If there are additional requirements for the product quality from the customer, such as gene cloning accuracy, refer to Annex F.

8.2 Synthetic genes

8.2.1 General

The quality control of synthetic genes includes the yield, purity and sequence.

8.2.2 Yield

The yield of synthesized genes can be determined by measuring the concentration of the re-suspended products with different approaches, such as ultraviolet spectrophotometry via determination of optical density at 260 nm (OD_{260}), or via fluorescence intensity quantitation using DNA binding fluorescent dyes. Quantitative PCR (qPCR) can be used as an additional validation method if the customer disputes results. Refer to B.2.

For the ultraviolet absorbance method, attention should be paid to the possible interference of unincorporated nucleotides and single-stranded nucleic acids.

8.2.3 Purity

The impurities in synthetic genes can include proteins, carbohydrates, guanidine, etc coming from the polymerase or ligase used for assembly, or from the host cell.

The ratio of OD at 260 nm versus 280 nm ($OD_{260/280}$) can be used to assess protein contamination. The value of $OD_{260/280}$ in the range of 1,8 to 2,0 indicates little protein contamination in the product. Refer to B.1.

The ratio of optical density at 260 nm versus 230 nm ($OD_{260/230}$) is used to assess other impurities such as carbohydrates, guanidine, etc. The $OD_{260/230}$ should be larger than 2,0 for normal plasmid and larger than 1,8 for low copy sample. Refer to B.1.

8.2.4 Sequence

The sequence of a synthetic gene product should be analysed by a DNA sequencing method such as Sanger or MPS to confirm whether or not the synthetic gene matches the sequence ordered by the customer. Each nucleotide should be assigned a numerical value (termed its "base quality score") that correlates to the predicted accuracy of the sequencing base call. Refer to B.4, Annex D and Annex E.

The following information should be provided to the customer per ordered sequence:

a) the target sequence file generated from the ordered sequence;

b) the alignment report of the synthetic genes product;

c) the sequence with the ordered (designed) sequence.

8.2.5 Integrity

Integrity reflects the degradation of synthetic double-stranded DNA. Refer to Annex C. After agarose gel or capillary electrophoresis, the molecular mass of the identified bands shall be consistent with the expectation.

Restriction digestion analysis also provides a way to confirm the integrity of synthetic double-stranded DNA. The manufacturer should select appropriate restriction enzymes and digestion methods to produce reliable digestion and differentiable band sizing.

8.2.6 Residual impurities analysis

Residual impurities in synthetic double-stranded DNA products such as endotoxin, host cell genomic DNA, host cell protein and residual mRNA should be assayed using appropriate methods upon customer request.

If the synthetic double-stranded DNA in a plasmid is used for transfection or the synthetic double-stranded DNA will be used for human or animal drug manufacturing, endotoxin in the product should be quantified using the Limulus amebocyte lysate (LAL) test method.

The levels of endotoxin should meet customer requirements and may generally fall within (0,1 to 1,0) EU/µg plasmid DNA, a level which is generally considered to have little negative effect on downstream applications. EU (endotoxin unit) is a unit assayed by the LAL test when testing for endotoxins.

NOTE 1 The residual genomic DNA in the product can be assayed by methods including hybridization and real time quantitative PCR upon request.

NOTE 2 If plasmids are prepared for use as, or in development of, drug substances/drug products, additional testing can be carried out.

Restriction enzyme analysis can be conducted by incubating the plasmid DNA with enzymes and analysing it on agarose gel or capillary electrophoresis. No other minor band should be observed; therefore, a recommended minimal amount is needed for reliable visualization.

8.2.7 Supercoiled plasmid

For synthetic double-stranded DNA in a plasmid, the percentage of supercoiled plasmid in total plasmid, which can affect the transfection efficiency and protein expression, can be assayed using agarose gel or capillary electrophoresis. The percentage of supercoiled plasmid in total plasmid can be determined at a condition defined by the customer.

8.2.8 Bioburden

Viable microorganisms contaminating a synthetic DNA product can be detected with a specific bioburden assay upon customer request.

8.2.9 Specific quality indicators for synthetic DNA libraries

The synthetic DNA library producer should provide evidence to the customer of the synthesis of the library by demonstrating conformity with the customer's specifications.

For synthetic DNA libraries, the following items shall be determined by DNA sequencing:

a) the library diversity;

b) the relative abundance of each DNA fragment;

c) the variants accuracy of each double-stranded DNA fragment.

Synthetic DNA libraries can be cloned into a specific vector to produce cloned libraries. A pre-determined number of clones (with statistical analysis) based on library size shall be selected randomly for sequence verification.

For an individual clone library, the sequence of the clone in the choice of vector should be verified by restriction digestion and DNA sequencing. NGS can also be applied to characterize the synthetic DNA library in pooled formats.

Although sequence, purity and impurities are important basic indicators of synthetic double-stranded DNA quality, the best indicator of quality is functionality in the downstream application of interest. If the customer finds a problem in the application, the quality shall be checked.

8.3 Synthetic genome

8.3.1 General

The quality control of a synthetic genome includes sequence. Sequence verification should be done to ensure the synthetic genome is correct.

8.3.2 Sequence

For synthetic genome products, sequence verification should be performed using appropriate DNA sequencing technology.

For synthetic genome, product variations should be less than 1 in 10 000 bp. If the mutation is located in a coding sequence, it should be a synonymous codon.

9 Delivered/synthesized material specifications

9.1 Main information

The delivered/synthesized material specifications for synthesized gene fragments, genes or genomes shall include, but are not limited to, the following:

a) amount of synthesized product;

b) DNA sequencing information: sequence alignment or statistical summary thereof;

c) purity information;

d) integrity information and restriction digestion map, if produced;

e) residual impurity information, if the customer requested this assay.

9.2 Other information

Other information that should be included in the final report of the delivered/synthesized material are:

a) producer information;

NOTE 1 Corresponding address, contact phone number, operator and time can be included.

b) product information;

NOTE 2 Name of product, length of product, sequence of the product can be included.

c) vector information, if the synthetic genes are cloned into a vector and the resulting plasmid is provided to the customer;

NOTE 3 The name of the vector, antibiotic resistance of the vector, sequence of the vector and plasmid map can be included.

d) host cell information, if the synthesized genes are cloned in a vector and replicated in a host cell, and are provided to the customer in the host cell containing the vector-based plasmid;

NOTE 4 The type of host cell, genotype and auxotype of the host cell, and culturing method of the host cell can be included.

e) product quality control information, including quality control results, methods and the analytical instruments used.

Annex A
(informative)

Example equipment and device list and their control requirements

An example of an equipment and devices list and their control requirements for the production and quality control of synthetic gene fragments, genes and genomes is given in Table A.1.

Table A.1 — List of equipment and devices and their control requirements

Types of equipment	Specifications	Intended use	Requirements	Inspection frequency
Temperature-controlled equipment (incubator, refrigerator, freezer, etc.)	—	Storage of reagents and samples	Temperature stability and homogeneity	At the time of installation, and then periodically[a]
			Temperature check	Real time
Pure water production system	—	Manufacturing/ quality control	Electric conductivity check	Periodically[a]
pH meter	—	Reagent preparation	Adjustment using at least two kinds of buffer solutions of proper quality	At the point of use
Weighing device	—	Reagent preparation	Confirmation of zero point and read check using standard mass	At the point of use
Pipettor or pipette	—	Dispensing of reagents	Calibration and check for accuracy of pipetting amount	Periodically[a]
Centrifuge machine	—	Centrifugation	Operate normally without abnormal noise	At the point of use
Oligonucleotide synthesizer	Devices that can perform synthesis of oligonucleotides	Oligonucleotide synthesis	Check for valves and flow volume of reagents	Periodically[a]
Electrophoresis equipment	Devices that can perform agarose gel electrophoresis under native conditions and detecting devices such as an ultraviolet lamp	Confirmation of genes lengths	Ability for separation and detection of appropriate molecular mass markers	Periodically[a]
Capillary electrophoresis system	Devices that can separate DNA/RNA or protein by mass under native conditions and are equipped with detectors and recorders that can sum up peak area	Confirmation of genes lengths	Ability for separation and detection of appropriate molecular mass markers	Periodically[a]
Automated pipetting system	Liquid handler, etc.	Dispensing of reagents and synthesizing genes, etc.	Check for accuracy of pipetting amount	Periodically[a]

[a] Inspection frequencies are defined for ensuring their fitness for their purpose, by considering the normal method of use and frequency (see ISO 9001 or standards related to the quality system of the organization).

Table A.1 *(continued)*

Types of equipment	Specifications	Intended use	Requirements	Inspection frequency
Dryer	Devices that are usable for the purpose of drying synthetic genes in solution	Drying of synthetic genes	Check for drying time	At the point of use
Ultraviolet spectrophotometer	Single beam or double beam spectrophotometers	Yield measurement of synthetic genes	Built-in performance check, or calibration by use of optical filter	Periodically[a]
Fluorescence spectrophotometer or fluorescence microplate reader	Single beam or double beam spectrophotometers	Yield measurement of synthetic genes	Built-in performance check, or calibration by use of optical filter	Periodically[a]
PCR machine	Devices that can carry out polymerase chain reaction	Genes synthesis and yield measurement	Temperature stability and homogeneity	Periodically[a]
DNA sequencer	Devices that can read out the sequences of synthetic genes and plasmids	Confirmation of DNA sequences	Performance test and calibration of core components (machine system, fluid system, optical system, etc.)	Periodically[a]

[a] Inspection frequencies are defined for ensuring their fitness for their purpose by considering the normal method of use and frequency (see ISO 9001 or standards related to the quality system of the organization).

Annex B
(informative)

Exemplary quality measurement methods

B.1 Test methods of synthetic nucleic acid purity

The ratio of optical density at 260 nm versus 280 nm ($OD_{260/280}$) is used to assess protein contamination coming from the polymerase or ligase used for assembly, or from the host cells. The ratio of optical density at 260 nm versus 230 nm ($OD_{260/230}$) is used to assess other impurities such as carbohydrates, guanidine, etc. The optical density is determined by a conventional or micro-volume ultraviolet spectrophotometer. A micro-volume ultraviolet spectrophotometer allows the determination to be quickly and easily run.

The pH value and ion concentration of the synthetic gene solution will affect the determination. Small changes in the pH value of the solution will cause the $OD_{260/280}$ to vary. Only in a certain pH value and lower ion concentration condition (e.g. pH 8,0, 10 mM Tris-HCl) is the result accurate. At the same time, the sample solution should be diluted to the appropriate concentration to make the optical density in the range of 0,1 to 1,0.

$OD_{260/280}$ in the range of 1,8 to 2 and $OD_{260/230}$ in the range of 2 to 2,2 is acceptable for DNA purity. If the purity ratio is significantly higher than expected, it is best to review the spectral profile as a primary means of troubleshooting.

B.2 Test methods of synthetic nucleic acid yield

The yield of gene fragments, genes or genomes products can be measured by different approaches including ultraviolet spectrophotometry method by the determination of optical density at 260 nm (OD_{260}), fluorescent quantitation with DNA binding fluorescent dyes and fluorometer, qPCR or digital PCR method.

For the ultraviolet spectrophotometry method, the absorbance of synthetic gene solution at 260 nm is determined according to the guide of the ultraviolet spectrophotometer. Then, the concentration of nucleic acid in the sample is calculated by using the Beer-Lambert equation and is reported in terms of mass/volume units (i.e. ng/µl). Linear range of the equipment is also established, and a high concentration solution should be diluted to fit in the linear range.

For the fluorescent method, add the aqueous solution of fluorescent nucleic acid stain for double-stranded DNA (dsDNA) into the cuvette or microlitre plate containing synthetic gene solution, then mix well and incubate for 2 min to 5 min at room temperature, protected from light. After incubation, measure the sample fluorescence using a spectrofluorometer or fluorescence microplate reader at standard fluorescein wavelengths. To ensure that the sample readings remain in the detection range of the fluorometer, the instrument's gain should be set appropriately so that the sample containing the highest DNA concentration yields fluorescence intensity near the fluorometer's maximum. To minimize photo bleaching effects, keep the time for fluorescence measurement constant for all samples. The use of RNase A/RNase T1 with S1 nuclease will eliminate all single-stranded nucleic acids and ensure that the entire sample fluorescence is due to dsDNA. Measure the fluorescence of the sample using instrument parameters that correspond to those used when generating the standard curve. Subtract the fluorescence value of the reagent blank from that of each of the samples. Determine the DNA concentration of the synthetic gene sample from the standard curve generated in the DNA standard curve. For the standard curve, bacteriophage lambda or calf thymus DNA is commonly used. The assay may be repeated using a different dilution of the sample to confirm the quantitation results.

For the qPCR method, the amount of targeted synthetic genes can be quantified by using specific primers of the corresponding genes and a calibration curve. The calibration curve can be constructed using independent standard samples with specified concentrations (e.g. copies/µl) that have the same or similar matrix as the test samples. The upstream processing of synthetic genes and the quantity, integrity and

purity of the samples should be determined to ensure that they are sufficiently pure, concentrated and free of components which inhibit or enhance the downstream qPCR reactions.

For the digital PCR method, the amount of targeted synthetic genes can be quantified without the use of a calibration curve. The dPCR mix containing a test solution is randomly distributed into discrete partitions of nominally equivalent volume such that some partitions contain no nucleic acid template and others contain one or more template copies. The partitions are thermally cycled to end-point and then read to determine the fraction of partitions with a positive reaction. Poisson statistics should be used to estimate the copy number of the target synthetic genes.

B.3 Measurement of synthetic nucleic acid integrity

Agarose gel electrophoresis including chip-based electrophoresis and capillary-based electrophoresis can be used to measure the synthetic nucleic acid integrity.

For agarose gel electrophoresis, the recommended values of agarose concentrations for DNA identification are given in Tables B.1 to B.2.

Table B.1 — Recommended agarose concentrations for different DNA length in electrophoresis

DNA length bp	Agarose concentration %
250 to 1 000	1,5 to 2
1 000 to 5 000	1,5
5 000 to 10 000	1

Table B.2 — Recommended agarose concentrations for different DNA length in pulse field electrophoresis

DNA length Mpb	Agarose concentration %
0,01 to 1	1,5
1 to 5	0,8 to 1
5 to 10	0,5 to 0,8

B.4 Sequencing

Sanger sequencing and MPS should be used for sequence verification, depending on the comprehensive consideration of the synthetic DNA producers. Usually, Sanger sequencing is used for short genes, and MPS is used for long genes. For a difficult template which can be challenging for MPS, Sanger sequencing is better suited to verify sequence identity.

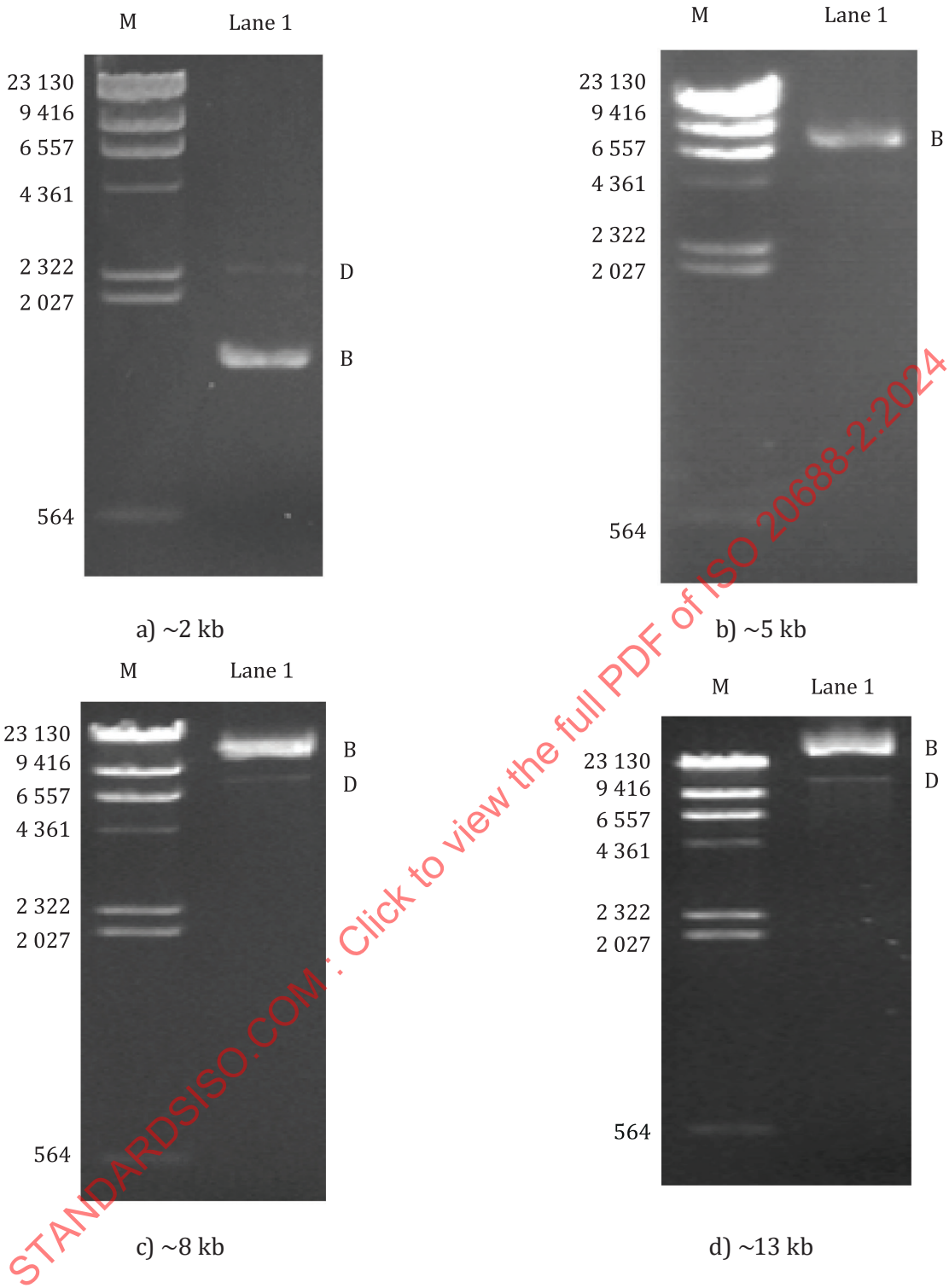
Annex C
(informative)

Electrophoretogram

If the synthetic gene fragments or genes are cloned into a plasmid vector, agarose gel electrophoresis can be used to measure its configurations and integrity.

In the process of plasmid extraction, the plasmid DNA strand can be broken due to mechanical force, pH, reagent, etc. The extracted plasmid usually exists in various kinds of configurations, including supercoiled plasmid DNA, linear plasmid DNA and open loop plasmid DNA.

Figure C.1 shows an electrophoretogram in 1 % agarose gel of plasmid DNA cloning different length synthetic genes of approximately 2 kb, 5 kb, 8 kb and 13 kb.



Key

M Lambda DNA/Hind III Marker

a) The brighter band (B)is superhelix plasmid DNA, and the darker band (D) is open loop plasmid DNA.

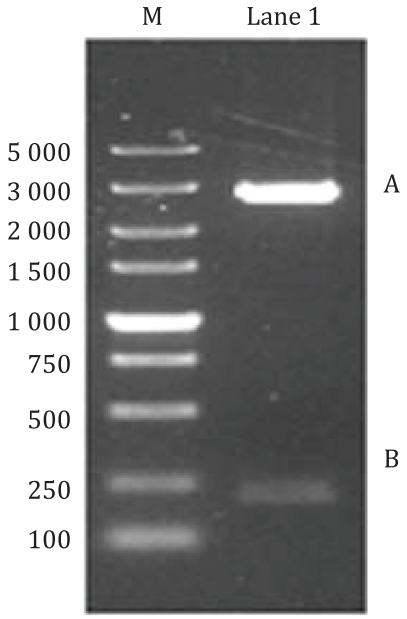
b) The bright band (B) is open loop plasmid DNA.

c) The brighter band (B) is open loop plasmid DNA, and the darker band (D) is linear plasmid DNA.

d) The brighter band (B) is open loop plasmid DNA, and the darker band (D) is linear plasmid DNA.

Figure C.1 — Electrophoretogram of plasmid DNA (1 % agarose gel)

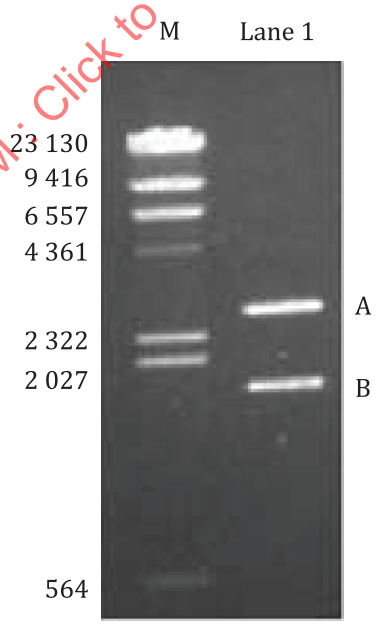
Agarose gel electrophoresis can also be used to analyse the base pair length of synthetic genes in plasmid vector after restriction digestion, as shown in Figure C.2, Figure C.3, and Figure C.4, which have different base pair lengths in different size vectors.



Key

A vector DNA
B synthetic gene
M Lambda DNA/Hind III Marker

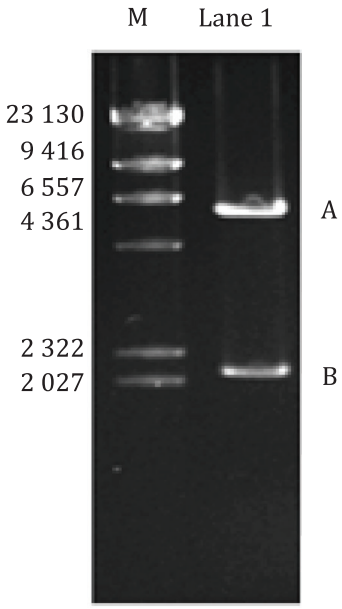
Figure C.2 — Restriction digestion analysis of synthetic gene of 246 bp in 3,1 kb vector (1 % agarose gel)



Key

A vector DNA
B synthetic gene
M Lambda DNA/Hind III Marker

Figure C.3 — Restriction digestion analysis of synthetic gene of 1,7 kb in 2,7 kb vector (1 % agarose gel)



Key

A synthetic gene

B vector DNA

M lambda DNA/Hind III Marker

Figure C.4 — Restriction digestion analysis of synthetic gene of 6,1 kb in 2,1 kb vector (1 % agarose gel)

Annex D
(informative)

Sanger sequencing

A widely used DNA sequencing method for synthetic genes is Sanger sequencing.

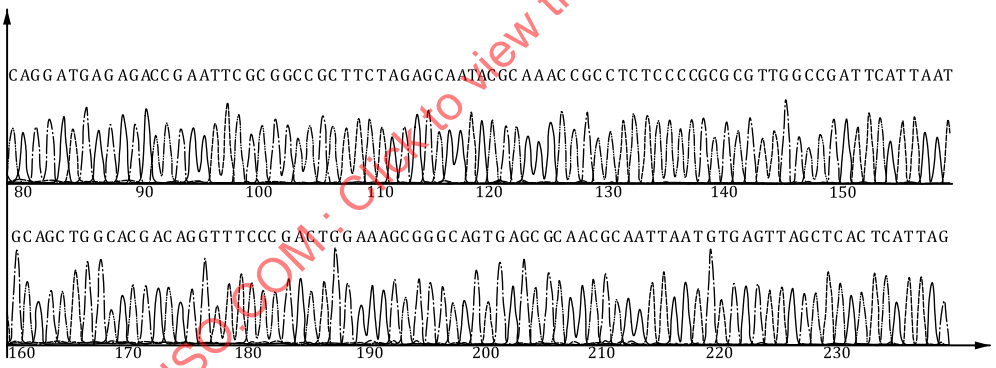
The synthetic gene is used as a template to produce a DNA ladder of fragments, each varying in length by one base, through randomly terminating DNA polymerase amplification by incorporating dideoxynucleotides.

In the commercial system developed by Applied Biosystems, specific dideoxynucleotides (ddATP, ddTTP, ddCTP and ddGTP) are each labelled with one type of fluorescence signal and incorporated into the end of each ladder DNA fragment. These fluorescently labelled ladder DNA fragments are then separated by capillary array electrophoresis (CAE). A fluorescence detector is then used to read the fluorescent signal. Fluorescence over time in each channel is used to provide a base case the resulting nucleotide read.

Figure D.1 illustrates the verification of a GFY gene by Sanger sequencing. Each peak represents a nucleotide signal. The result shows that the sequence of synthesized GFY gene is 100 % correct.

Sequence

> GFY-Gene
CAGGATGAGAGACCGAATTCGCGGCCGCTTCT AGAGCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGAT
TCATTAATGCAGCTGGCACGACAGGTTTCCCG ACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGT
TAGCTCACTCATTAG



Key

_____	A
—·—·—	C
—·—	G
– – –	T

Figure D.1 — Verification of a GFY gene by Sanger sequencing