

Ref.: 2014-08108-01  
CWCh/PNi

## APPENDIX 6

Non-CCI version

Does not include confidential commercial information

**Serine protease from *Nocardiopsis prasina*  
produced by a genetically modified strain of *Bacillus  
licheniformis***

**Novozymes A/S**  
August 29<sup>th</sup> 2014

Regulatory Affairs

Novozymes A/S  
Krogshøjvej 36  
DK-2880 Bagsværd  
Denmark

Telephone:  
+45 4446 0000  
Telefax:  
+45 4446 4647

E-mail:  
Internet:  
[www.novozymes.com](http://www.novozymes.com)

CVR number:  
10 00 71 27

## Appendix 6

### non-CCI version

***Elements in Appendix 6 that are to be treated as confidential commercial information (CCI) are marked with a red box in the CCI version. The confidential information has been removed from the non-CCI version Grey colour has been used for the applied redactions.***

### Documentation regarding the production strain

1. Detailed description of the construction of the genetically modified production strain
2. Description of general methods.
3. Annotated DNA sequence *amyL* region
4. Annotated DNA sequence *xyIA* region
5. Genetic stability of the production strain (Southern blot)

## Appendix 6.1

### Detailed description of the construction of the genetically modified production strain

#### 6.1.1. The host organism

##### Taxonomy

The parental strain Ca63 is a natural isolate and the taxonomic classification is as followed:

Name:	<i>Bacillus licheniformis</i>
Phylum:	Firmicutes
Class:	Bacilli
Order:	<i>Bacillales</i>
Genus	<i>Bacillus</i>
Species:	<i>licheniformis</i>

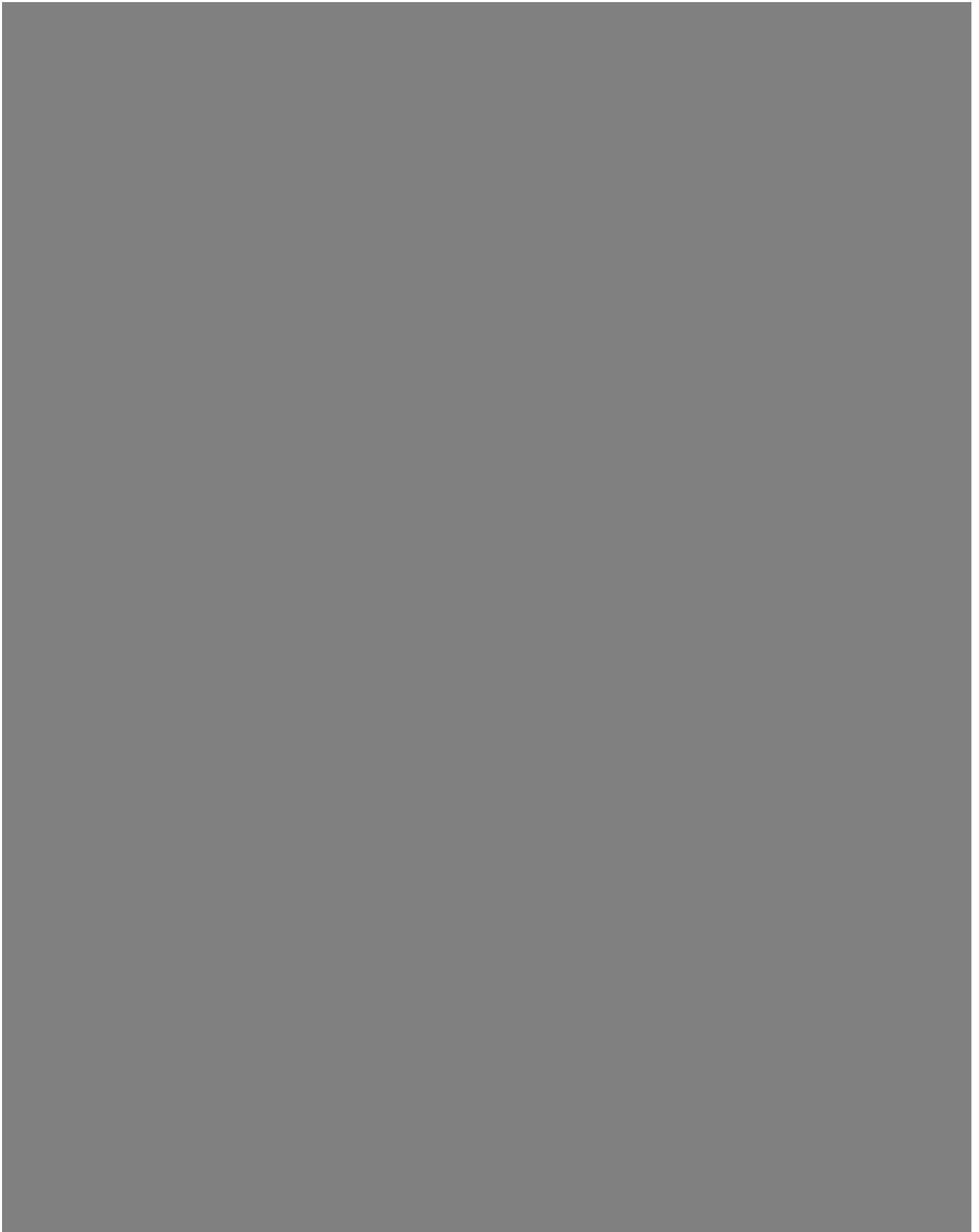
The classification was confirmed by Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH and deposited as DSM 9552 (equals to ATCC 9789).



##### Genetic modifications

Host strain *B. licheniformis* SJ6370 was constructed from parental strain Ca63 [REDACTED] (Fig. 1). The modifications are briefly described below. The general methods used for engineering the strain are described in Appendix 6.2. The loci modified during the GM steps leading from parental strain Ca63 to host strain SJ6370 are listed in Table 1.

Locus	Function
<i>aprL</i>	Alkaline protease
<i>amyL</i>	Alpha-amylase
<i>mprL</i>	Glu-specific protease
<i>xylA</i>	Xylose isomerase

**Table 1:** Loci modified in GM steps during construction of host strain SJ6370





No antibiotic marker genes were introduced into the host strain as a result of these genetic modifications.

### 6.1.2 Origin and donor of vector and inserts

#### The enzyme gene

The serine protease is encoded by a tandem gene construct inserted at the *amyL* and *xylA* loci in the chromosome. Each tandem gene construct consists of two genes

namely the serine protease from *Nocardiopsis prasina*

In total, four copies of the gene encoding the are present in the production strain and these are hereafter collectively referred to as

#### Promoter

During construction of the host strain SJ6370, a triple tandem promoter construct is inserted at the *amyL* and *xylA* locus, respectively. This triple tandem promoter construct consists of three promoter fragments, originating from three different donor organisms.

The three promoters are:

#### Terminator

The transcriptional terminator (*amyL* term) inserted downstream of the serine protease construct at both the *amyL* and *xylA* locus is from the *Bacillus licheniformis* DSM 9552.

#### Vector/insert

Vectors used are composed of elements from plasmids pUB110 (Gryczan *et al*, 1978) and pE194 (Horinouchi, S. and Weisblum, B., 1982).

### 6.1.2. Introduced genetic sequence

#### Construction of production strain from host strain SJ6370

The production strain *B. licheniformis* S10-34zEK4 was constructed from the host strain *B. licheniformis* SJ6370 using the double homologous recombination strategy explained in the general methods (Appendix 6.2). Two different plasmids containing the tandem gene construct described in the section above were used to insert a total of four copies of the downstream of the modified triple tandem promoter present at the *amyL* and *xylA* loci. Further, the ribosome binding site (RBS) in front of

■ was replaced with a non-functional RBS version leading to an abolished production of this background protein.

Fig. 2 presents the pedigree leading from host strain SJ6370 to the final production strain, including the genetic change introduced in each step and the plasmids used in the process.

The methods used during strain construction are described in Appendix 6.2.









To confirm the number and position of the four [REDACTED] copies that have been inserted in the production strain S10-34zEK4, Southern blot analysis was performed on strains in the pedigree leading from host strain to the production strain with a probe specific for [REDACTED] genes. Bands of the expected sizes were obtained, confirming that all four [REDACTED] copies are inserted at the target loci in a correct manner.

To verify that the intended modifications have been introduced in *amyL* and *xylA*, the DNA sequences of these loci were determined in the final production strain. The annotated sequence of the *amyL* locus containing the serine protease genes is given in Appendix 6.3 whereas the annotated sequence of the *xylA* locus containing the serine protease genes is given in Appendix 6.4. The sequences were determined as described in the general methods (Appendix 6.2) and were found to match the expected sequences.

To demonstrate that no ARM genes from these used plasmids were present in the final production strain, Southern blot analysis was performed on chromosomal DNA from this strain using different probes designed to recognize relevant target sequences (ARMs) of these vectors. The results showed absence of ARM genes in the final production strain.

### **6.1.3. Description of the production organism**

The chromosome of the final production strain has been modified by recombinant DNA techniques at five different positions relative to the non-recombinant strain Si3. These positions are

- A) The position of the gene encoding an alkaline protease where a deletion was introduced.
- B) The position of the gene encoding a glu-specific protease, where a deletion was introduced.
- C) The position of the gene encoding a background protein, [REDACTED], where the ribosome binding site was modified.
- D) The position of the gene encoding the alpha-amylase, *amyL*, where the expression cassette with the tandem serine protease gene construct was inserted.
- E) The position of the gene encoding the xylose isomerase, *xylA*, where the expression cassette with the tandem serine protease gene construct was inserted.

The resulting strain was subsequently subjected to classical mutagenesis and a production strain S10-34zEK4 giving a high yield was selected.

#### Identity and taxonomy of production organism

The production strain for the serine protease is a *Bacillus licheniformis* carrying four genes encoding the serine protease from *Nocardiopsis prasina*. The strain is derived from a sporulation deficient *Bacillus licheniformis*, as described above.

#### Genetic stability and mobilization and conjugation capability

The inserted recombinant DNA is genetically stable during fermentation, as the inserted DNA is integrated into the chromosome.

The genetic stability of the production strain was tested at large-scale fermentation. The strain stability during fermentation was analyzed by Southern blotting. No instability of the strain was observed (Appendix 6.5).

As the insert is chromosomally integrated and lacks a functional origin of replication, it cannot be transferred by conjugation to other organisms, nor can fragments replicate autonomously.

#### Antibiotic resistance gene

No functional antibiotic resistance genes were left in the strain as a result of the genetic modifications.

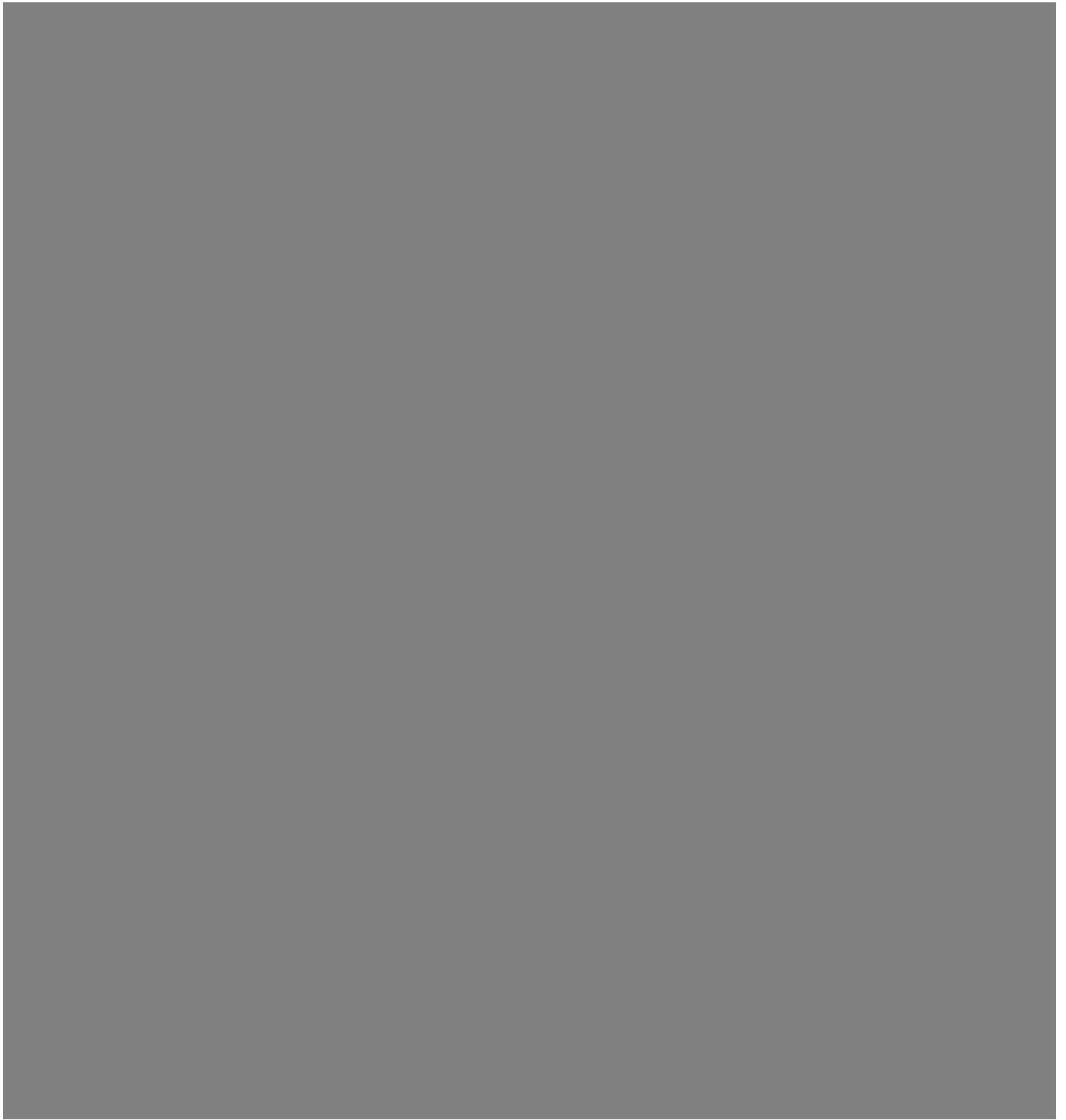
## Appendix 6.2

### General description of methods













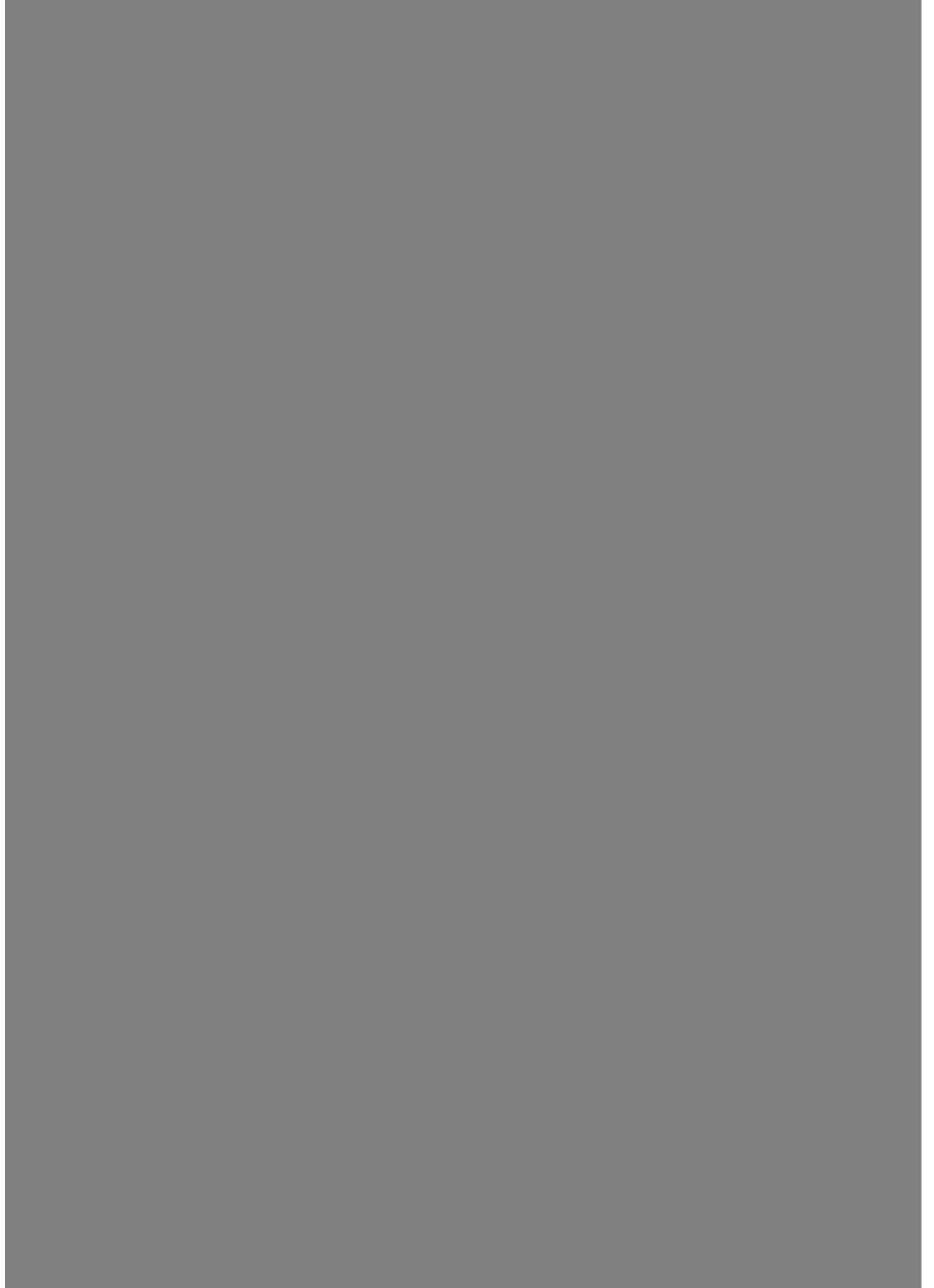


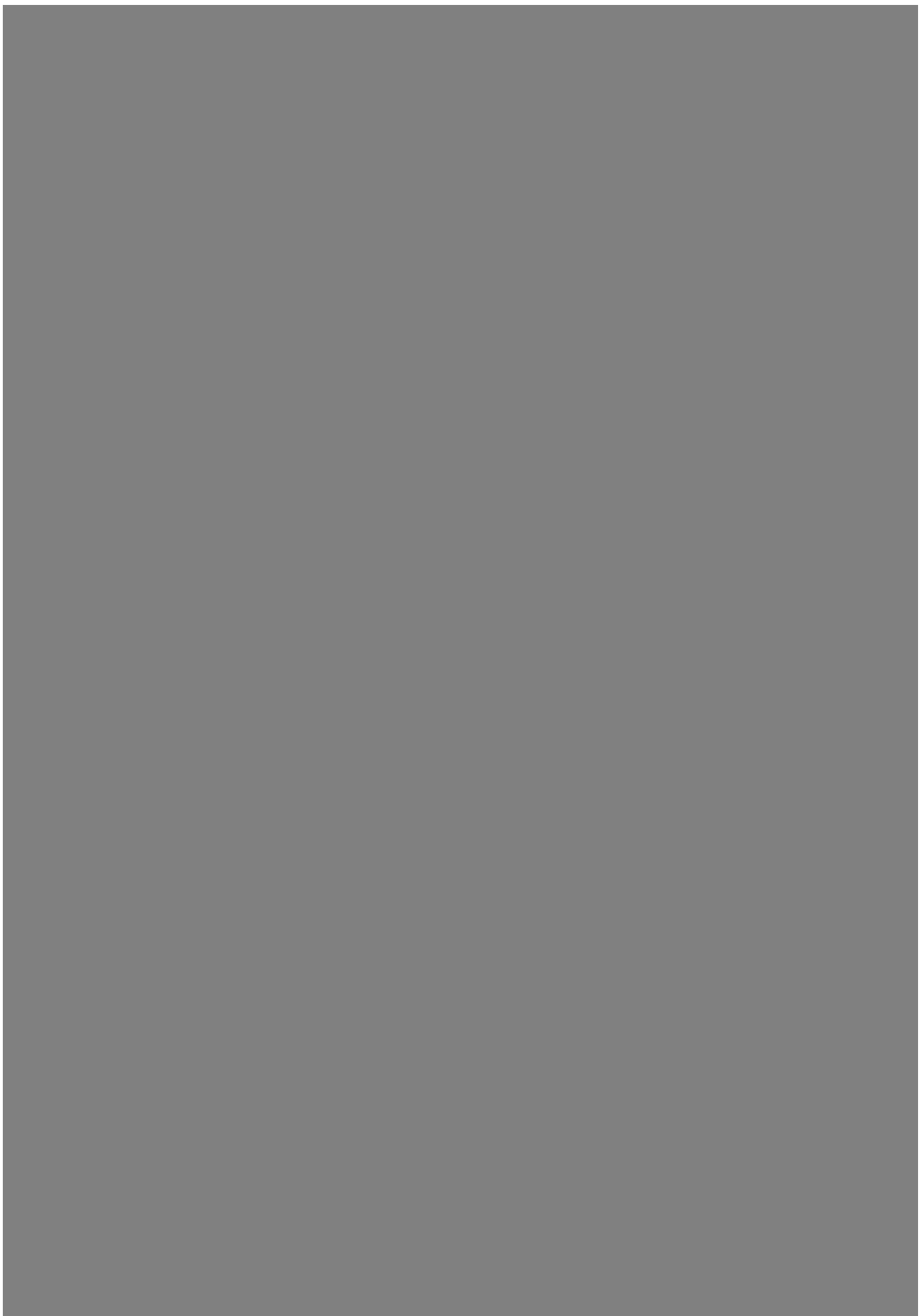
## Appendix 6.3

### Characterization of the *amyL* locus in the production strain

To verify that the modification had been introduced in *amyL* as intended, the DNA sequence of the locus in the production strain was determined as described under General methods (Appendix 6.2). The overview of the *amyL* locus and the annotated sequence in the production strain can be found below.

Annotated sequence of the *amyL* locus in the production strain:









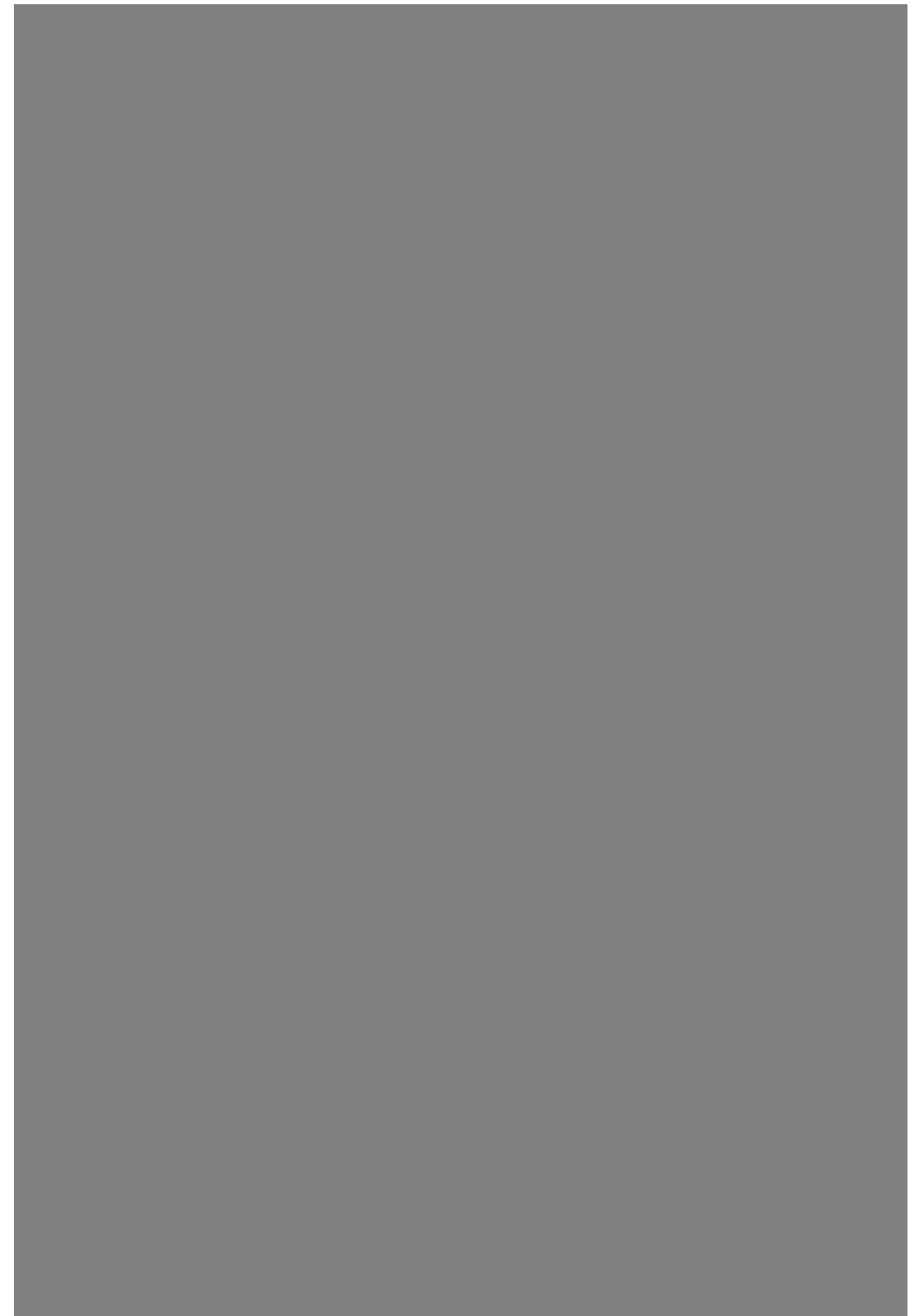
## Appendix 6.4

### Characterization of the *xyfA* locus in the production strain

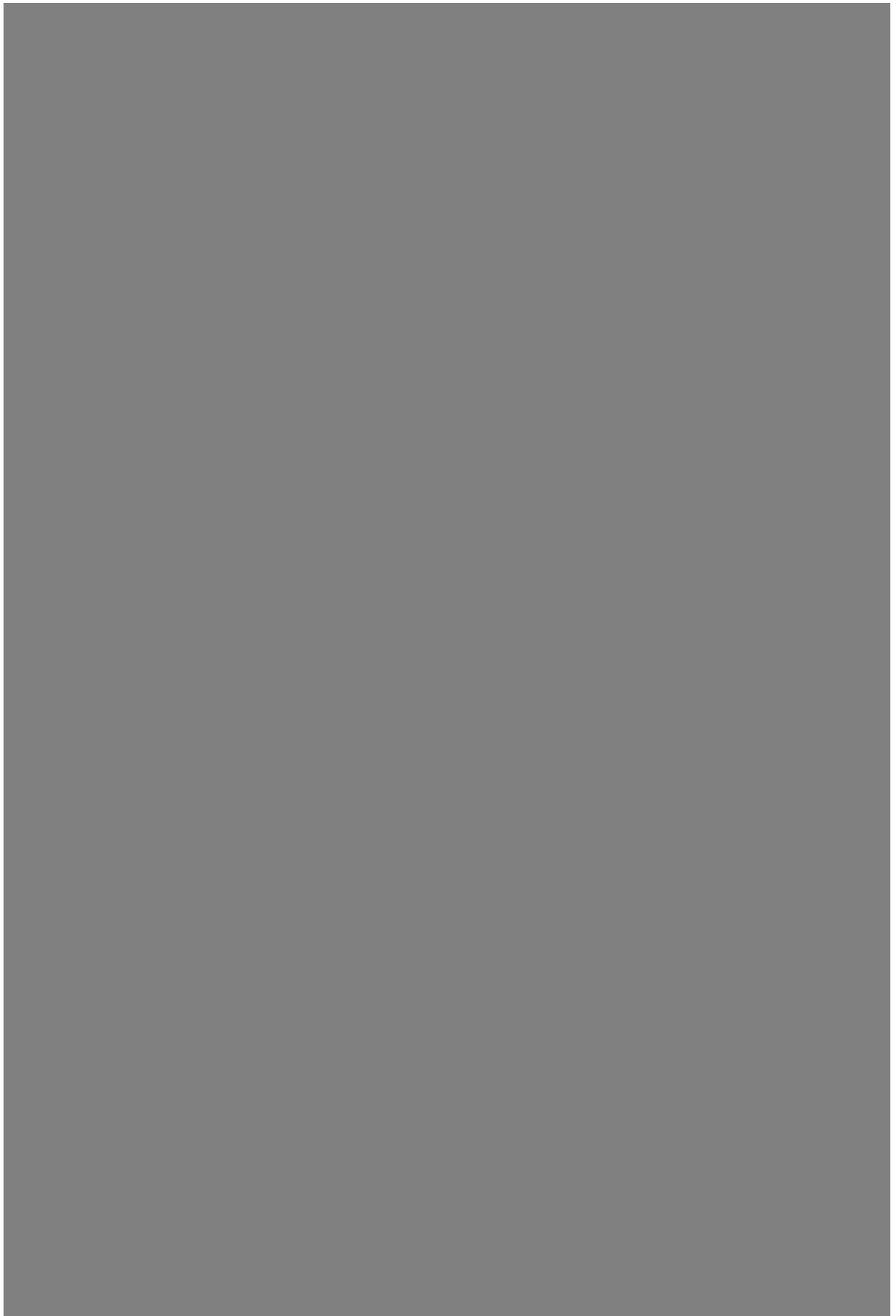
To verify that the modification had been introduced in *xyfA* as intended, the DNA sequence of the locus in the production strain was determined as described in the general methods (Appendix 6.2). The overview of the *xyfA* locus and the annotated sequence in the production strain can be found below.

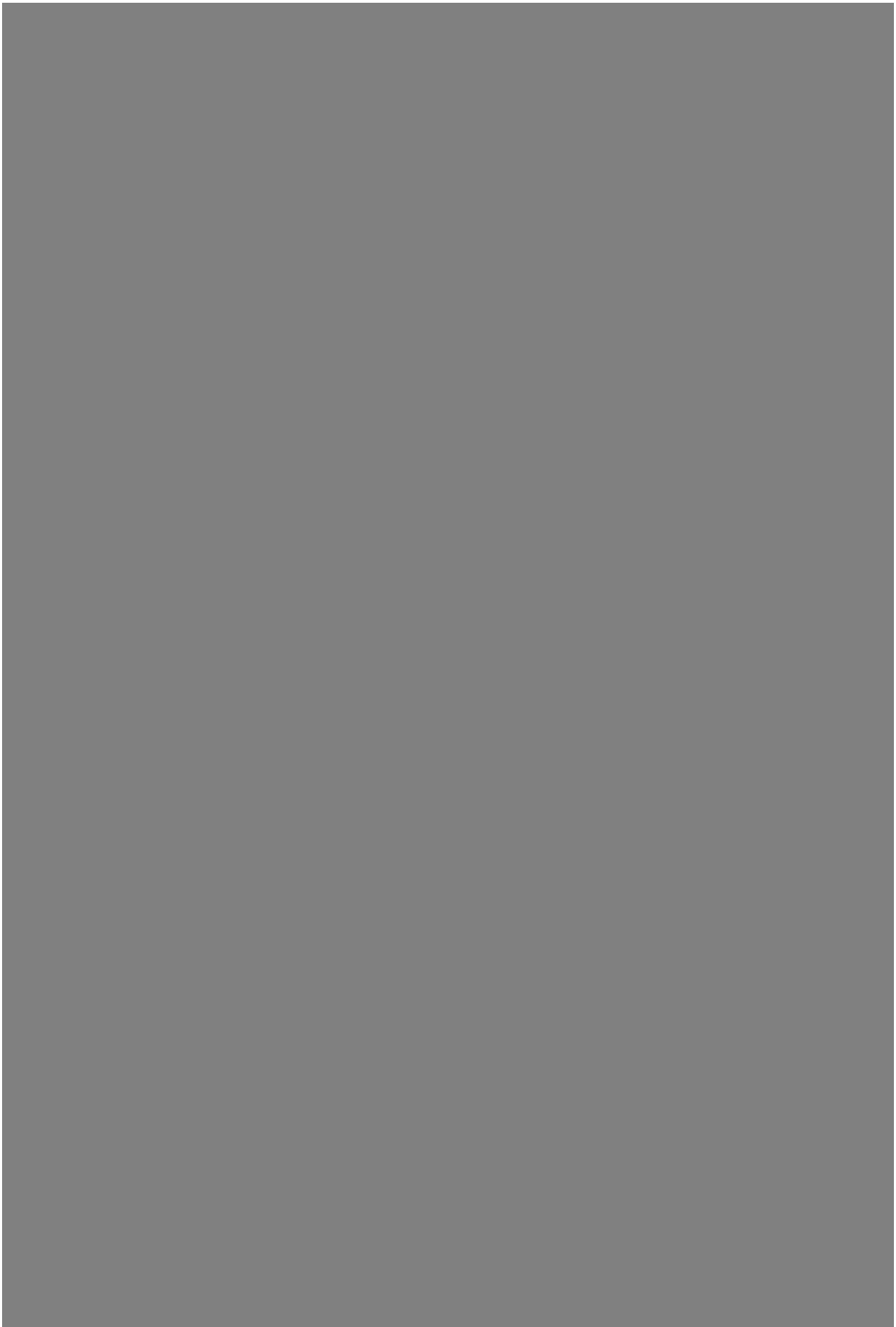




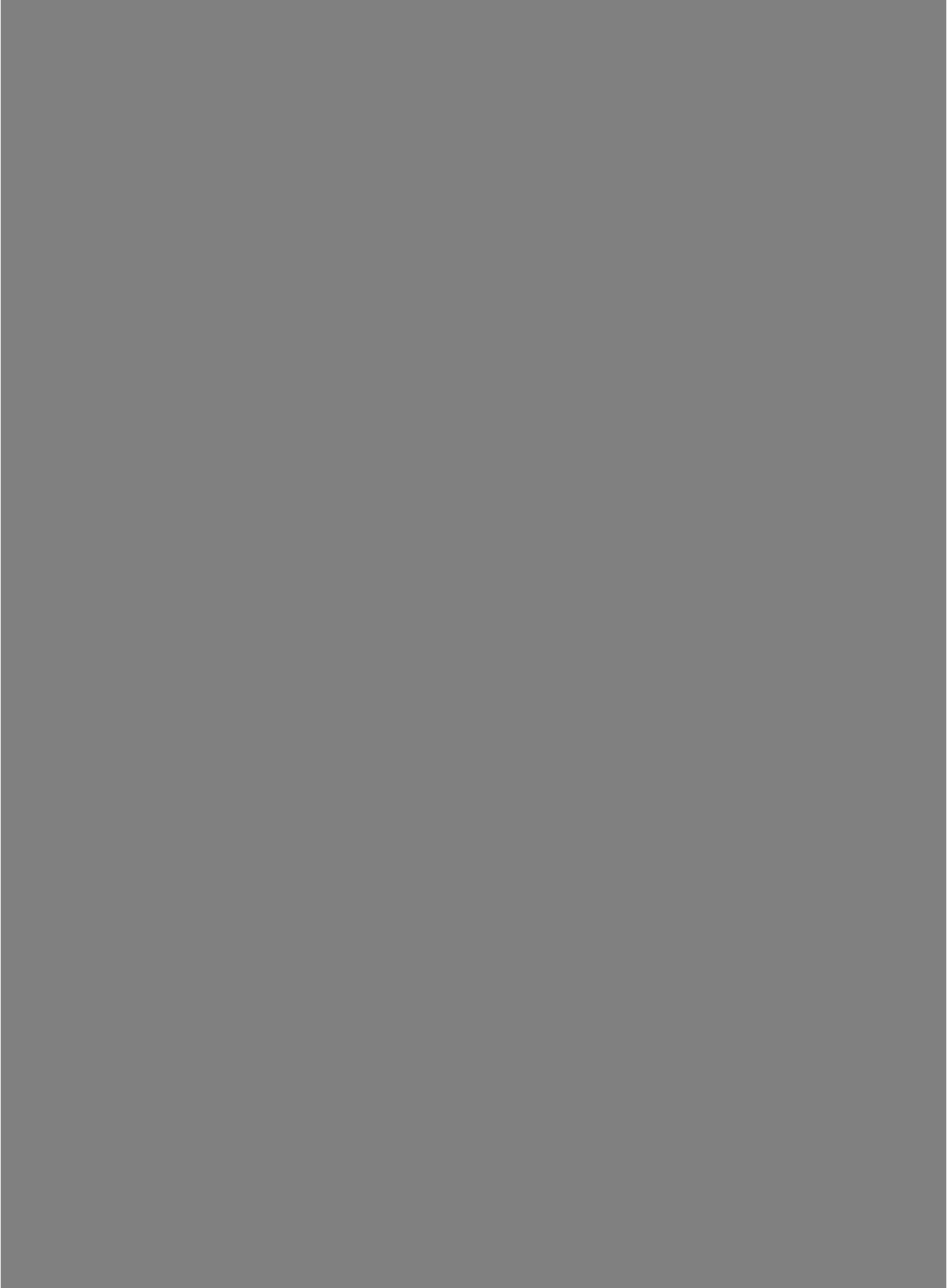












The presented results verify the presence of 2 copies of the [REDACTED] tandem gene construct (i.e., a total of 4 gene copies of [REDACTED], one copy present at each of the targeted loci, *amyL* and *xylA*. Southern blot analysis showed identical band pattern between the production strain and the samples derived from end of production. Thus, it is concluded that the production strain is genetically stable.

#### Method for Southern blot analysis

Genomic DNA was purified from the samples and [REDACTED]. DNA fragments were separated on an 0.75% agarose gel and transferred to a nylon membrane. Hybridization was performed using a [REDACTED] (Fig. 2).

## References for Appendix 6

Gryczan TJ, Contente S, Dubnau D (1978) Characterization of *Staphylococcus aureus* Plasmids Introduced by Transformation into *Bacillus subtilis*. *Journal of Bacteriology*, 134 (1), 318-329.

Horinouchi S, Weisblum B (1982) Nucleotide Sequence and Functional Map of pE194, a Plasmid that Specifies Inducible Resistance to Macrolide, Lincosamide, and Streptogramin Type B Antibiotics. *Journal of Bacteriology*, 150 (2), 804-814.