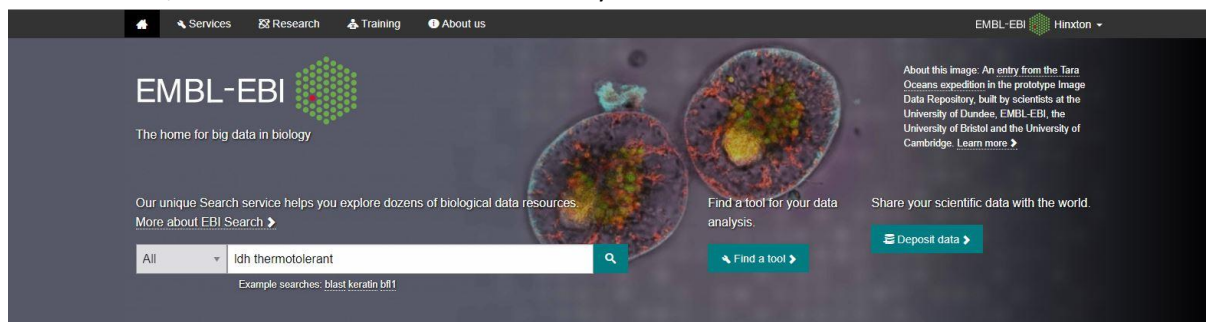


Bioinformatics

LDH sequence

1. For creating lactate from pyruvate we needed an enzyme, called lactate-dehydrogenase (LDH). Therefore we used the database of the European Bioinformatics Institute (www.ebi.ac.uk) to find an appropriate enzyme. Since the *Methylococcus capsulatus* is a thermotolerant organism, we thought it would be worth searching for a thermotolerant LDH as well, so we made the search with the keywords “ldh thermotolerant”.



The screenshot shows the EMBL-EBI homepage. At the top, there are navigation links for Services, Research, Training, and About us. The main header features the EMBL-EBI logo and the tagline "The home for big data in biology". Below this, there is a search bar containing the query "ldh thermotolerant". To the right of the search bar, there are buttons for "Find a tool" and "Deposit data". A small text box on the right side of the page provides information about the image, mentioning the Tara Oceans expedition and the University of Dundee, EMBL-EBI, the University of Bristol, and the University of Cambridge.

Explore EMBL-EBI and our mission

The European Bioinformatics Institute (EMBL-EBI) shares data from [life science experiments](#), performs [basic research](#) in computational biology and offers an extensive [user training](#) programme, supporting researchers in academia and [industry](#). We are part of [EMBL](#), Europe's flagship laboratory for the life sciences. [More about EMBL-EBI and our impact](#)

Services

We provide freely available data and bioinformatics services to all facets of the scientific community

Research

We contribute to the advancement of biology through basic investigator-driven research

Training

We provide advanced bioinformatics training to scientists at all levels

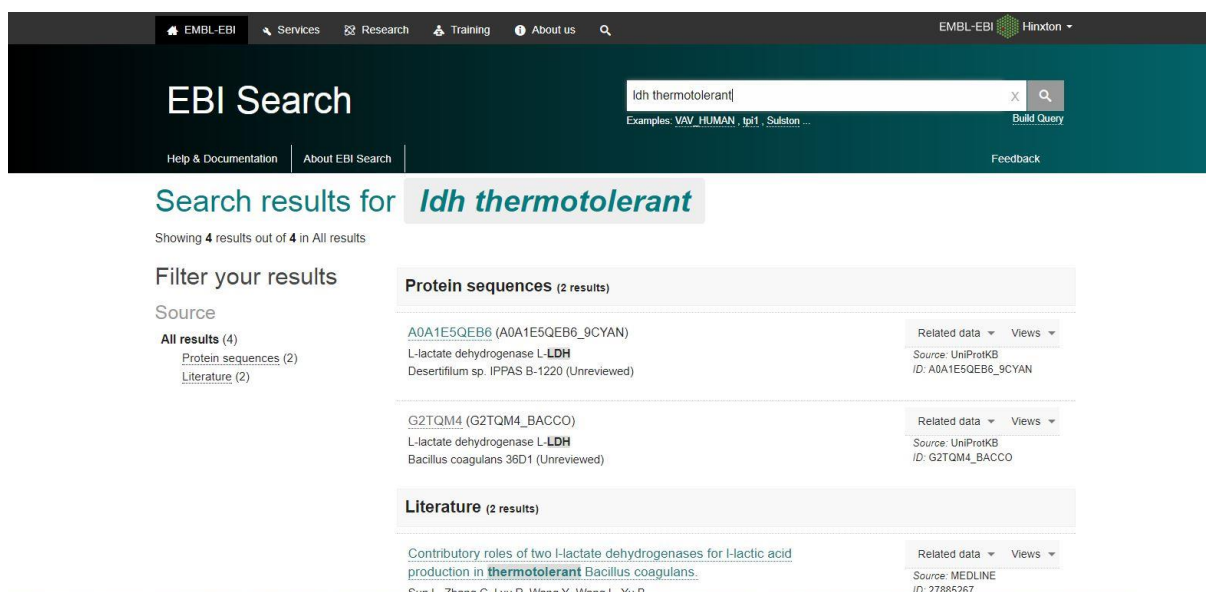
Industry

We help disseminate cutting-edge technologies to industry

ELIXIR

We support, as an ELIXIR node, the coordination of biological data provision throughout Europe

2. Searching for the keywords “ldh thermotolerant” two results were shown. One of them was the protein sequence from the *Bacillus coagulans* 36D1 bacterium and the other one from the algae species *Desertifilum*. For using it in a bacterium we chose the bacterium enzyme.



The screenshot shows the EBI Search results page for the query "ldh thermotolerant". The search bar at the top contains the query and a search button. Below the search bar, there are links for "Help & Documentation", "About EBI Search", and "Feedback". The main heading is "Search results for ldh thermotolerant". Below this, it says "Showing 4 results out of 4 in All results". There is a "Filter your results" section with a "Source" filter. Under "All results (4)", there are two "Protein sequences (2)" and two "Literature (2)".

Source	Protein sequences (2)	Literature (2)
Protein sequences (2)	<p>A0A1E5QEB6 (A0A1E5QEB6_9CYAN) L-lactate dehydrogenase L-LDH Desertifilum sp. IPPAS B-1220 (Unreviewed)</p> <p>G2TQM4 (G2TQM4_BACCO) L-lactate dehydrogenase L-LDH Bacillus coagulans 36D1 (Unreviewed)</p>	<p>Contributory roles of two L-lactate dehydrogenases for L-lactic acid production in thermotolerant <i>Bacillus coagulans</i>. Sun L, Zhang C, Lyu P, Wang Y, Wang L, Yu B</p>
Literature (2)		

- In the UniProt database (www.uniprot.org) we could find valuable information (e.g. function, names and taxonomy, structure, sequence etc.) about the chosen enzyme.

UniProtKB - G2TQM4 (G2TQM4_BACCO)

Protein | L-lactate dehydrogenase
Gene | ldh
Organism | *Bacillus coagulans* 36D1
Status | Unreviewed - Annotation score: ●●●○○ - Protein inferred from homology¹

Function
Catalytic activity¹
 (S)-lactate + NAD⁺ = pyruvate + NADH. UniRule annotation -
Pathway¹: pyruvate fermentation to lactate
 This protein is involved in step 1 of the subpathway that synthesizes (S)-lactate from pyruvate. UniRule annotation -
 Proteins known to be involved in this subpathway in this organism are:
 step 1: L-lactate dehydrogenase (**ldh**)
 This subpathway is part of the pathway pyruvate fermentation to lactate, which is itself part of Fermentation.
 View all proteins of this organism that are known to be involved in the subpathway that synthesizes (S)-lactate from pyruvate, the pathway pyruvate fermentation to lactate and in Fermentation.

Feature key	Position(s)	Description	Actions	Graphical view	Length
Binding site ¹	90	Substrate UniRule annotation -			1
Binding site ¹	122	NAD or substrate UniRule annotation -			1
Binding site ¹	153	Substrate UniRule annotation -			1
Active site ¹	177	Proton acceptor UniRule annotation -			1

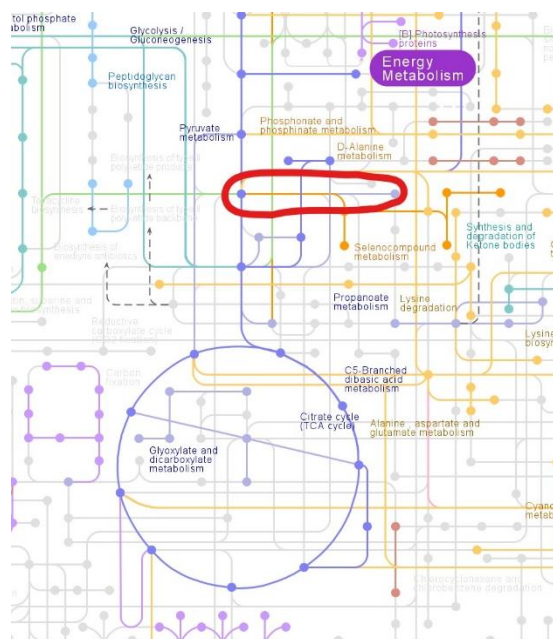
- For finding the gene sequence of the LDH, we visited the website of the KEGG PATHWAY Database (<http://www.genome.jp/kegg/pathway.html>), where we looked for the metabolic pathway of *Bacillus coagulans*.

KEGG Metabolic pathways - Bacillus coagulans 36D1

[Pathway menu | Organism menu | Pathway entry | Hide module list | User data mapping | Image (png) file]

Bacillus coagulans 36D1 100%

- In the figure we had to find the reaction of the lactate-pyruvate converting for the enzyme details.



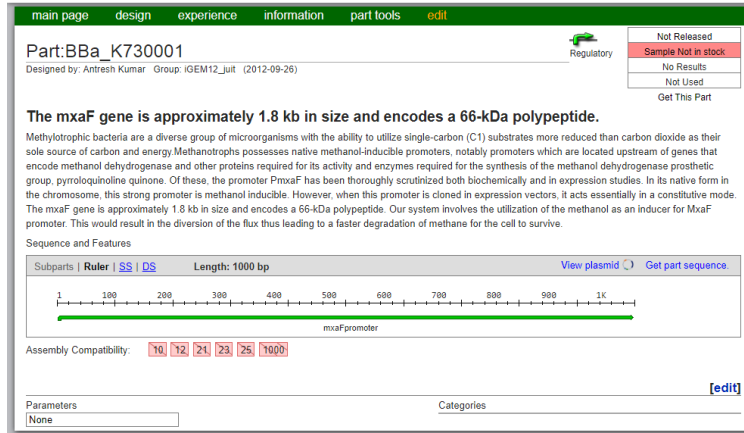
6. Opening the enzyme's datasheet we realized that two enzymes are in connection with this reaction. We chose the L-lactate-dehydrogenase because it was smaller in size and had the usual beginning ATG bases.

KEGG **Bacillus coagulans 36D1: Bcoa_0653** Help

Entry	Bcoa_0653 CDS T01628	All links Ontology (3) KEGG BRITE (3) Pathway (6) KEGG PATHWAY (8) Chemical substance (9) KEGG COMPOUND (9) Chemical reaction (4) KEGG ENZYME (1) KEGG REACTION (3) Genome (1) KEGG GENOME (1) Gene (3) KEGG ORTHOLOGY (1) NCBI-PROTEINID (1) OC (1) Protein sequence (1) UniProt (1) Protein domain (6) Pfam (6) All databases (35) Download RDF
Definition	(GenBank) L-lactate dehydrogenase	
KO	K00016 L-lactate dehydrogenase [EC:1.1.1.27]	
Organism	bag Bacillus coagulans 36D1	
Pathway	bag00010 Glycolysis / Gluconeogenesis bag00270 Cysteine and methionine metabolism bag00620 Pyruvate metabolism bag00640 Propanoate metabolism bag01100 Metabolic pathways bag01110 Biosynthesis of secondary metabolites bag01120 Microbial metabolism in diverse environments bag01130 Biosynthesis of antibiotics	
Brite	KEGG Orthology (KO) [BR:bag00001] Metabolism Carbohydrate metabolism 00010 Glycolysis / Gluconeogenesis Bcoa_0653 00620 Pyruvate metabolism Bcoa_0653 00640 Propanoate metabolism Bcoa_0653 Amino acid metabolism 00270 Cysteine and methionine metabolism Bcoa_0653 Enzymes [BR:bag01000] 1. Oxidoreductases 1.1 Acting on the CH-OH group of donors 1.1.1 With NAD+ or NADP+ as acceptor 1.1.1.27 L-lactate dehydrogenase Bcoa_0653 Exosome [BR:bag04147] Exosomal proteins Exosomal proteins of epithelial cells Bcoa_0653 Exosomal proteins of breast milk Bcoa_0653 BRITE Hierarchy	
SSDB	Ortholog Paralog Gene cluster GFIT	
Motif	Pfam: Ldh_1_N Ldh_1_C UDPG_MGDP_dh_N ApbA 3HCDH_N TrkA_N Motif	
Other DBs	NCBI-ProteinID: AEO99872 UniProt: G2TQM4	
Position	complement(693722..694660) Genome map	
AA seq	312 aa AA seq DB search MKKVNRIAVVGTGAVGTSYCYAMINQGVAEELVLIDINEAKAEGEAMDNLHGLPFAPTPT RWVKGDYSDCGTADLVVITAGSPQKPGETRLDLVAKNAKIFKGMKIMSDFNGIFLVA SNPVDILTYVTWKESGLPKHEVIGSGVLD SARLRNSLSAHFGIDPRNVHAAIIEHGDT ELPVWHSHTTIGYDTIESYLQKGTIDOKLDDIFVNTRDAAYHIERKGFATFYIGIGMSLR ITRAILNNSVLTVSFAFLEGGYVNSDVIYIGVPAVINRQGVREVVEIELNDKEQEFSHS VKVLKETMAPVL	
NT seq	939 nt NT seq +upstream 0 nt +downstream 0 nt atgaaaaggctcaatcgattgcaagtggggaacgggtgcagttggtcaagttactgc tacgccatgattaatcagggtgtgcaagaagagctgttttaacgatattaacgaagca aaagcagaaggggaagccatggacctgaaccacggcctgccattgcccctacggcagacc cgctttggaaggcgattatccgattgcccagcctgcccgatctgttgcattacggca ggttcccccaaaaacgggcgaaacaagcctgatctgtgcccataaacgcaaaaatt tttaaggcatgattaagagcatatggacagcggcttaacgggattttctgtgccc agcaaccgggtgacattttgacatattgacttggaaagagtcggcctgcccgaagaa catgttaccggttcggccacagtgctgactccggcgtctccgcaacttttaagcgc cactcgggaattgaccgcgcaatgctccgcaatfatcggcgaacacggcgcacag gaactccggtttggaccatacaacgatcggttatgacaccattgaagcctatctgcaa aagggaaccattgaccatacaacattagatgattttgcaacacgagagatggcgt taccatatacgaagaaaaggccacatttaccgcatcgggatgctctgaccgg atcacaagagcagctcgaacaalgaaaaacagttttgacagctcctgcttttgaa ggccagtacggaacacagcagatgtgacatgggttctccgcttataaccgcaagc gtccgtaagtggtgaaatcgagcgaagcaaaaagcaggaacaatttagccattct gttaaaagtataaaaagcagatggaccctgattgtaa	

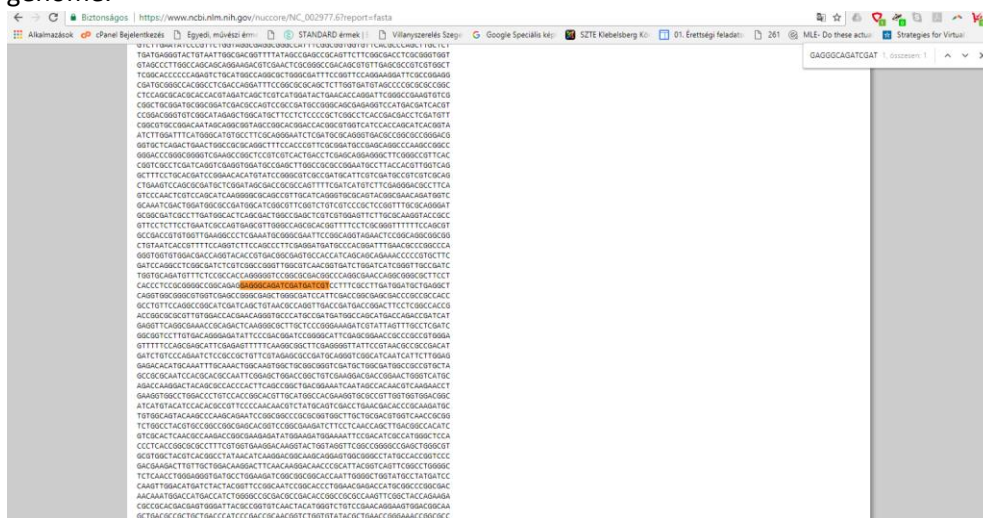
Assembling the final construct

- It was also essential to find an appropriate promoter with high efficiency to enable the gene to work. Therefore, we looked for a promoter which can be found originally in *Methylococcus capsulatus*. The former iGEM team, iGEM12_juit found a nearly 1.8 kb long sequence which contained a promoter but they could not determine the exact location of the promoter.



The screenshot shows the iGEM parts database page for Part:BBa_K730001. The page title is "Part:BBa_K730001" and it was designed by Antrish Kumar from the iGEM12_juit group. The part is a regulatory element, 1000 bp long, and is not released. The description states that the *mxhF* gene is approximately 1.8 kb in size and encodes a 66-kDa polypeptide. It is a strong promoter in methanol-inducible systems. The page includes a sequence viewer with a ruler showing the 1000 bp length and assembly compatibility options (10, 12, 24, 25, 1000). The parameters are set to "None" and the categories are empty.

- We did a search to find the 1.8 kb long sequence in *Methylococcus capsulatus*' complete genome.



The screenshot shows a BLAST search result for a 1.8 kb sequence in the *Methylococcus capsulatus* genome. The search was performed on the NCBI website. The top hit is "Methylococcus capsulatus strain ATCC 29618", with a score of 1000.00 and an identity of 100%. The alignment shows a perfect match between the query sequence and the database sequence. The query sequence is 1800 bp long, and the database sequence is also 1800 bp long.

NCBI Reference Sequence: NC_002977.6
 GenBank FASTA

Methylococcus capsulatus str. Bath, complete genome

Search Results

Label	From	To	Strand
GAGGGGATCGATGATCGT	822734	822753	Positive

Page 1 of 1 | Displaying Search Results 1 - 1 of 1

Related information

- Assembly
- BioProject
- BioSample
- Components (Core)
- Full text in PMC

- This sequence contained not only the nucleotide sequence between two genes (the *moxY* and the *moxX*) but did contain a partial part of each of the two genes.

GenBank FASTA

NC_002977.6:823K..823K (47bp) Find: [] Tools [] Tracks []

Genes

STS Markers

Repeat region

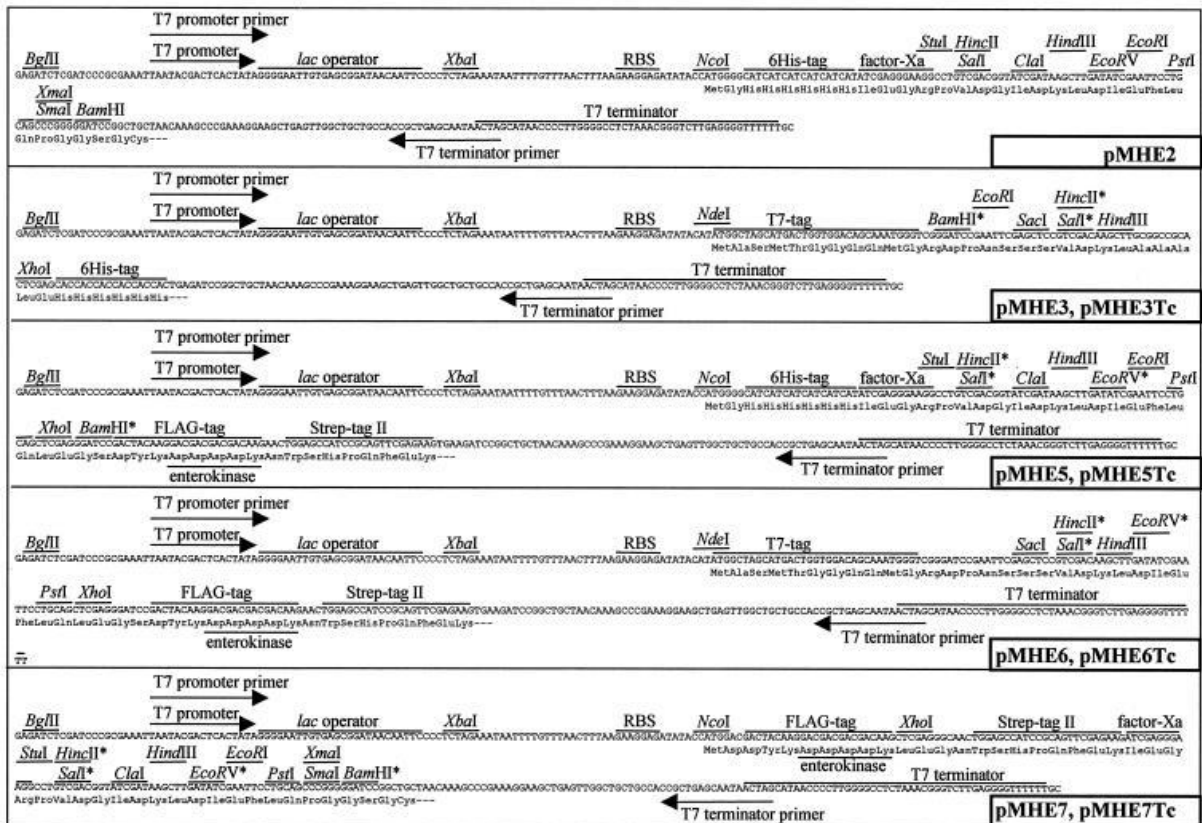
MCA_RS03845
 Gene: MCA_RS03845
 Location: complement(821,671..822,990)
 Length: 1,320
 [Qualifiers]
 old_locus_tag: MCA0778
 CDS: WP_010960105.1
 Title: methanol utilization control sensor protein moxY
 Location: complement(821,671..822,990)
 [Length]
 Span: 1,320
 Product: 439
 [Qualifiers]
 inference: COORDINATES: similar to AA sequence:RefSeq:WP_010960105.1
 Download: WP_010960105.1

Range: 823040..823047
 Zoom On Range
 Zoom To Sequence
 Modify Range
 Add New Panel On Range
 Set New Marker For Selection
 BLAST Search (Selection)
 Primer BLAST (Selection)
 Download FASTA (Selection)
 Download GenBank Flat File (Selection)

WP_010960106.1
 CDS: WP_010960106.1
 Title: PQQ-dependent dehydrogenase, methanol/ethanol family
 Location: 823,278..825,083
 [Length]
 Span: 1,806
 Product: 601
 [Qualifiers]
 inference: COORDINATES: similar to AA sequence:RefSeq:WP_010960106.1
 Download: WP_010960106.1
 Links & Tools
 BLAST Genomic: NC_002977.6 (823,278..825,083)
 BLAST Protein: WP_010960106.1
 BLINK Results: WP_010960106.1
 FASTA View: NC_002977.6 (823,278..825,083), WP_010960106.1
 GenBank View: NC_002977.6 (823,278..825,083), WP_010960106.1
 Graphical View: WP_010960106.1

- The putative *moxY* or *moxX* used in this study did only contain the nucleotide sequence between the two genes, which must contain the promoter of either the *moxY* or the *moxX* gene. Unfortunately, the orientation of the promoter is not known because the *moxY* and the *moxX* genes are in different directions, therefore we could not determine neither the exact orientation of the promoter nor the exact sequence but we managed to approach the exact sequence of the promoter and apply it in such a way that the orientation was not needed to know.

- Having the sequence of the LDH gene and an appropriate promoter, only a usable vector was needed to transfer the gene and the promoter into *Methylococcus capsulatus*. An article wrote about the vectors pMHE2, pMHE3, pMHE5, pMHE6 and pMHE7 (figure). We chose the pMHE5 and pMHE7 vectors because they were available for us in short time.



Figure

6. The LDH gene was intended to ligate between the NcoI and EcoRV restriction sites, therefore, it was supplied with an NcoI restriction site and an addition AGTCAGTC nucleotide sequence before and after the NcoI restriction site, in order to minimize the possible damage made by the restriction enzymes (figure).

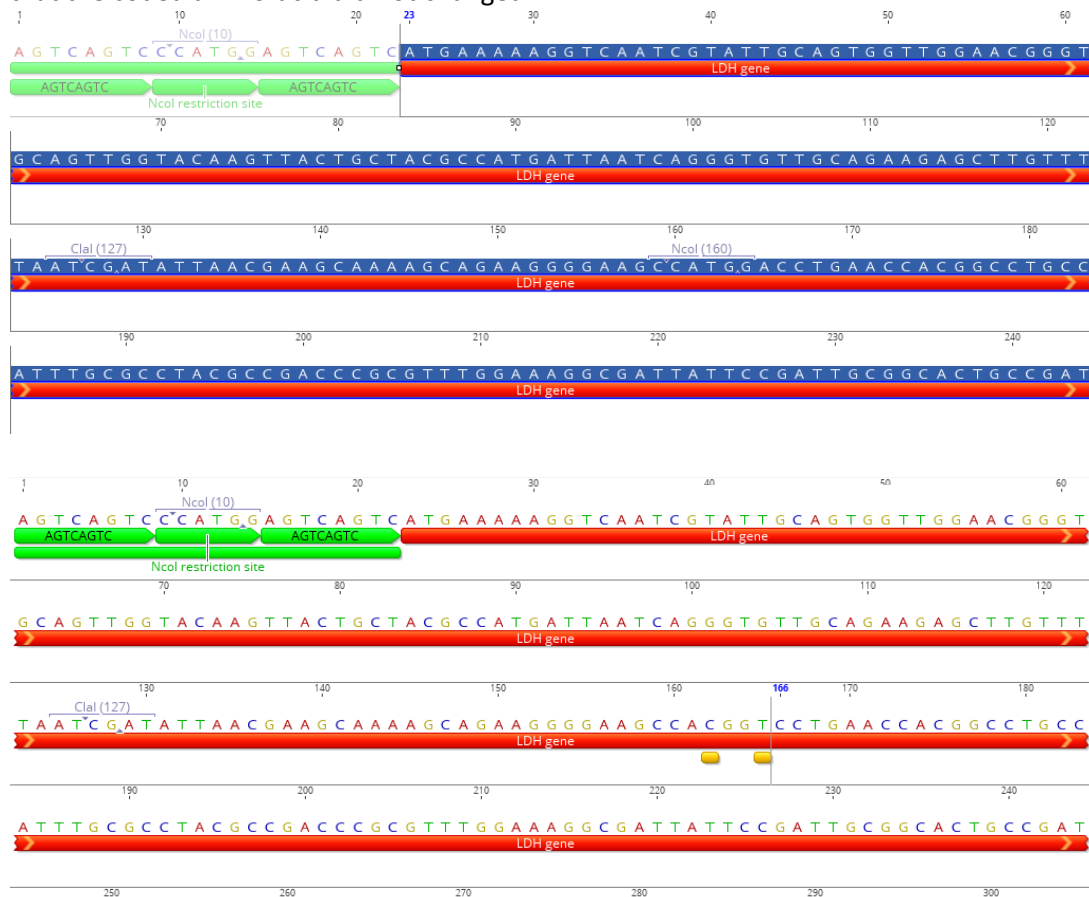


Figure: The basic structure of the synthesized gene

1. Addition AGTCAGTC bases
2. NcoI restriction site
3. LDH gene
4. EcoRV restriction site

7.

The LDH gene contained an NcoI restriction site in itself, therefore, we replaced the bases so that the coded amino-acid did not changed.



8. The promoter was intended to ligate between the BglII restriction sites. This could enable the promoter to ligate in both orientations, randomly. The promoter was supplied with BglII restriction sites at both ends and addition AGTCAGTC nucleotides before and after each added AGTCAGTC nucleotides (figure).

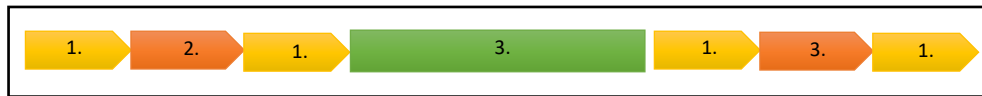


Figure: The basic structure of the synthesized promoter

1. Addition AGTCAGTC bases
2. BglII restriction site
3. Promoter