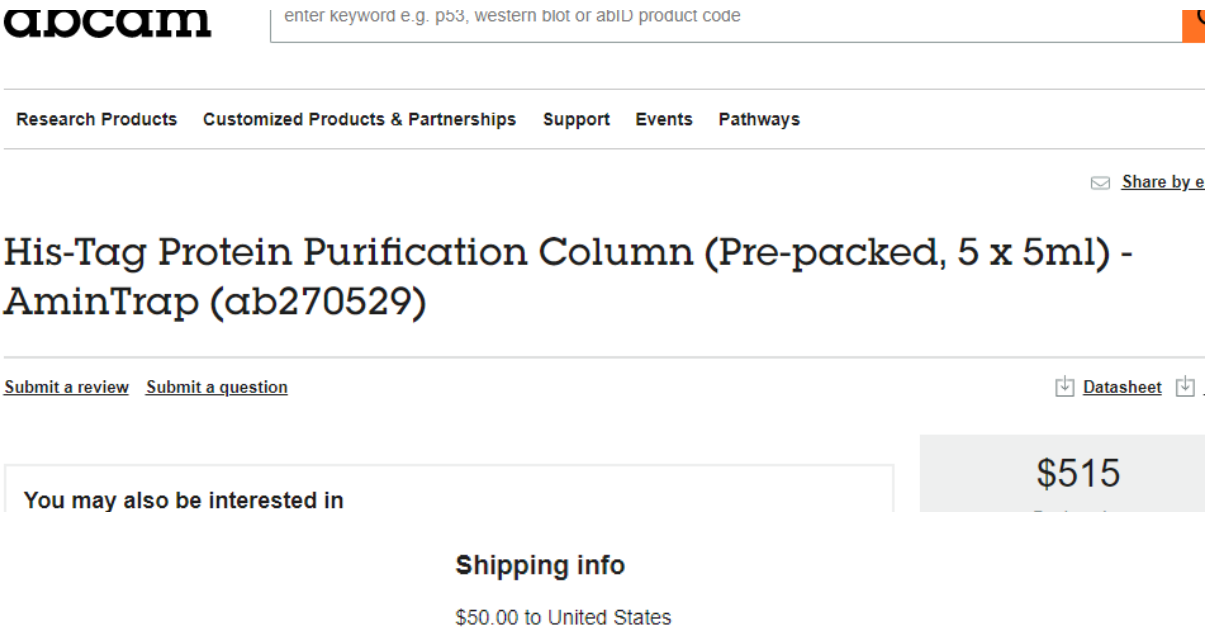
Can we significantly reduce the high costs of enzymes to make biology and medical research accessible and affordable ?

Abstract

Modern molecular biology approaches to research, health, biomaterials etc need devices and reagents. Reagents cost is usually higher upfront and in long term because the start quantity is usually higher than the needs for one experiment. More so, it's a recurring cost. Some reagents like enzymes have reduced shelf time, require expensive cold chain transportation. Usually shipping costs are high and many times prohibitive in low resource settings.
The pandemic demonstrated the supply chains are brittle and unreliable. During the first part of the pandemic it was almost impossible to obtain Reverse Transferase which is a must for research of any RNA virus.
Bioproduction consists normally of two parts where the first part is where the specific product is created and the second part is where the specific product is purified/extracted. The second part, downstream processing is as important as the first part and most of the time more difficult and expensive. While that is somehow counter-intuitive, keep in mind that production of proteins is using a natural process and usually once the proper plasmids have been transformed into a proper organism it's only the matter of feeding the organism and maintaining some parameters in a certain range. Downstream processing on the other hand, is an artificial process where the living organism membrane is lysed and we need to use some type of downstream purification that requires extra reagents and equipment. **The most commonly used process today is using affinity chromatography columns that contain Nickel beads that fuse a 6xHis tag to the product of interest**. Nickel columns are not inexpensive and can only be reused at most several times.

Here is an example not from the most expensive provider:



Column size: 1.6 x 2.5cm.

Binding capacity: >40mg His-tagged protein/ml medium

Particle size: 45-165µm.

Of course that is favorable to the companies that sell these columns so they would not be promoting an alternative that requires no further purchases.

There is an alternate solution where the **nickel columns and reagents are replaced with silica coated magnetic beads and extraction is done by using permanent magnets**. In principle, the silica coated magnetic beads are totally reusable. The key is finding a short fusion protein that would bind to silica and this is what this project is about.

Wouldn't be good to spend less time asking for funds and more time doing research?

Aims

Find a usable silica tag that will offer a reusable inexpensive alternative and pack it in a standard format to require minimum, simple editing. As we aim for standardization and simplification we should attempt to create one or more constructs as a standard synthetic bio part, otherwise known as a bio-brick https://en.wikipedia.org/wiki/BioBrick. We also aim to release this as an Open Source so we should identify a patent free sequence to be usable as an open source by any lab. As the sequence will be used in subsequent genetic processes like cloning we also need to make sure the sequence does not interfere with prospective protocols especially cloning method. We might need to create a software component to correctly identify sequences we want to avoid and suggest new sequences. Identify a suitable induction system candidate
Reduce the cost, and complexity of affinity purification

Specific aims

- Find an affinity tag for silica binding
- Identify it's copyright status. Perform a patent search to verify that is royalty free
- The tag has to be express-able in e-coli as e-coli is one of the most used production organisms
  The tag has to be under 20 AA to not have a significant effect on the product due to length or volume. Other problems, like hairpins should be eliminated
- Find a cleavable linker to allow for cleavage of the tag
- Create one or more constructs as a Standard synthetic bio part, otherwise known as a BioBrick https://en.wikipedia.org/wiki/BioBrick that can use a simplified cloning procedure called MoClo that does not require PCR (PCR reactions are prone to introduce mutations)
- Processing needed for MoClo (like domestication/optimization) should not affect the binding site(s). That requires us to identify and replace forbidden motifs mainly by substitution with silent mutations.
- Develop software to design and domesticate/optimize the tag sequence
- Identify a suitable induction system candidate that uses physical induction to reduce cost, reagent dependence and remove need for carcinogenic reagents like IPTG.
- Reduce the cost of affinity purification by at least 10-fold, reduce purification time to minutes, reduce the dependence on reagents and eliminate special shipping and storage requirements.

Method

Hunting for a patent free silica Tag

Literature research yields few hits that are in open source space. One candidate is described in

SI-4 leads to (Si4-x) developed by Naik et al (Journal of Nanoscience and Nanotechnology, Volume 2, Number 1, February 2002, pp. 95-100(6))" (4)

MSPHPHPRHHHT which is Si4-1 variant of 12 AAs

Naik used some computations to create a library and he developed Si-n group where Fig3. Si4-1 was the best but we also have Si3-8 and Si4-8 pretty high in mols/silica. In relative units Si4-1 was still the best Fig 2.
Naik filed numerous patent applications. Here is the last one:
https://www.lens.org/lens/bio/patseqfinder#/results/ae57bb4f-4d04-4294-9cb7-05282bda3c74?docTypeLocations=Grant-CLAIM&locations=grantInClaims&offset=0&limit=25
however none of them became patents for technical reasons.

This tag sequence originates in a marine animal. While we have several variants of the sequence. Si4-1 was the best but we also have Si3-8 and Si4-8 pretty high in mols/silica. In relative units Si4-1 was still the best Fig 2.

So we select:
MSPHPHPRHHHT which is the Si4-1 variant

Ideally we would also test Si3-8 and Si4-8 and Si3-3.

We know it binds to Silica and we know is patent free. Only testing will determine if it would show activity and would work well with product recovery.

This gene originated in diatoms, a type of algae. Interestingly enough this organism produces 20 to 30% of the air we breathe. https://diatoms.org/what-are-diatoms

More options.

There were some other candidates yet a closer look revealed they were under patent protection

For instance (6)

The Cleavable Linker aim

From FreeGenes row cleavage linker encoding for a short TEV_CUT on row 60

“TEV protease recognition linker, useful for protein purification. N terminal tags leave almost no linker behind after cleavage”

According to Removal of Affinity Tags with TEV Protease (7) “TEV protease recognizes the amino acid sequence ENLYFQ/G with high efficiency and cleaves between Q and G”

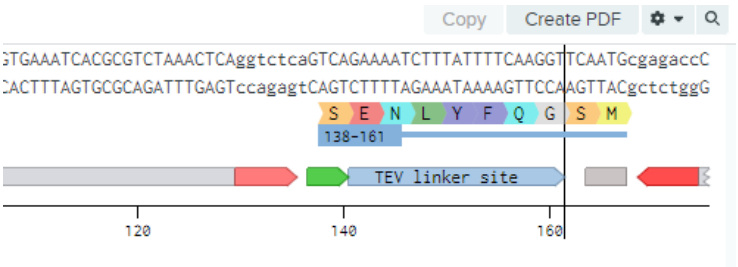
https://docs.google.com/spreadsheets/d/1uUS6TaIy0YuT8W7hdeonu0NSpAHinieyPcvgZRG1Nos/edit#gid=1530011293

Benchling: https://benchling.com/s/seq-Nm1fDcYOd0Abd8ottzj6 GAAAATCTTTATTTTCAAGGT which is the Naked cleavable peptide DNA sequence

For N terminal compound silica affinity tag+cleavage we will be using FreeGenes construct flanked between N2(GTCA) and C(AATG) 4 bases fusion sites. Here is the sequence that includes the MoClo Prefix and suffix and the flanking Bsa1

and an annotation with the naked cleavable peptide DNA sequence. Note an extra 2 bases (AG) added between the end of the cleave sequence and the MoClo suffix to stay in Frame with the next Biobrick that will be a CD.

Bsa1 recognition sites in Red, Moclo N2 prefix in Green and MoClo suffix(C) in gray, naked sequence annotated in Blue and an A and a G added to stay in frame with the next CD.



N2C_TEVcut =https://benchling.com/openbioeconomy/f/lib_JdqC4iAO-cleavage-sites/seq_LEmo6z39-n2c_ncut_tev/edit

There are some other iGem choices as per http://parts.igem.org/Protein_domains/Linker

Design of the standard BioBrick part aim

MoClo (Modular Cloning) based on Golden Gate (flanked by sites recognised by type IIS restriction enzymes) requires flanking the parts with specific short 4 nucleotide prefixes and suffixes (5) called fusion sites (2).

All parts are prefixed and suffixed by standard PhytoBrick (one implementation of BioBrick standards) by 4 nucleotides fusion sites that will determine the joining order during the assembly in a one pot reaction. We use fusion sites similar to to
https://docs.google.com/spreadsheets/d/1uUS6TaIy0YuT8W7hdeonu0NSpAHinieyPcvgZRG1Nos/edit#gid=807017147

BsaI, BpiI, and BsmBI are prohibited in this BioBrick release.

The tag can be placed at the N terminal or at the C terminal. Testing is necessary to determine the best placement.

We need to place the cleavable linker between the tag and the CDS of the gene of interest so cleavage will remove the tag.

Standard BioBricks parts needs to be flanked by two BsaI sites in opposing orientations as the ending is cut by Bsa1 scanning on the complementary strand sometimes called template so sequences ggtctcannnn and nnnntgagacc for the 5′ and 3 sites .

where BSA1 recognition site is:

Recognition sequence	Cut
5' GGTCTC 3' CCAGAG	5' ---GGTCTCN NNNN--- 3' 3' ---CCAGAGNNNNN --- 5'

Most importantly, since the silica binding tag and the linker have to be coexpressed with the gene of interests' CDS' all 3 must be in frame. So interceding sequences must be multiples of 3 nucleotides.

Option 1:

Order just the affinity tag as a BioBrick sequence and then assemble it later with the FreeGenes linker. This option is more laborious but will create an atomic BioBrick containing just the Silica affinity tagin case one will want not to use the linker.

N Terminal conceptual sequence:

some nucleotides, BSA1 recognition site, MoClo prefix tag for N term , Silica Tag , MoClo suffix tag for N Term, BSA1 recognition site on opposite strand, some nucleotides,

NNNNN **GGTCTCAN , CCAT , (domesticated converted MSPHPHPRHHHT to DNA) , NN GTCA, NGAGACC NNNNN**

where NN were introduced to keep the element starting at ATG in frame with what will be added after the suffix. We still need to make sure the codon created by NN and G is a viable codon for e-coli.

C Terminal, full conceptual sequence
BSA1 recognition site , MoClo prefix tag for C Term position , Silica Tag , Stop codon, MoClo suffix tag for C Term position , BSA1 recognition site

C Terminal, sequence

GGTCTCAN, CGGC *NN* , (domesticated converted MSPHPHPRHHHT to DNA) , Stop codon, NN GCTT, NGAGACC

Note: After digestion with Bsa1 the parts in Blue will be lost on the "top-coding" strand the 5' prime will have an overhang (CCAT) and the 3 prime (other strand) will have an overhang too

Use https://www.bioinformatics.org/sms2/rev_trans.html to convert AA seq to DNA for e-coli

MSPHPHPRHHHT →ATGAGCCCGCATCCGCATCCGCGCCATCATCATACC

Option 2: Order both fragments (affinity tag and linker as one sequence)

Proposed N Terminal conceptual sequence to order in case Option 2 is selected:

landing, BSA1 recognition site , MoClo prefix tag for N term, Silica Tag , MoClo suffix tag for N term, MoClo prefix tag 2 , Linker, MoClo suffix tag 2, BSA1 recognition site on opposite strand, landing for endonuclease

GGTCTCAN - CCAT *NN* (domesticated converted MSPHPHPRHHHT to DNA) - GTCA *NN*-GGAGGTGGCTCTGGCGGTGGATCAGGTGGAGGCTCT - AATG GAAAATCTTTATTTTCAAGGT NGAGACC

C Terminal is also possible barring that the previous CDS was stripped of the stop codon.

Note: After digestion with Bsa1 the parts in Blue will be lost on the "top-coding" strand the 5' prime will have an overhang (CCAT) and the 3 prime (other strand) will have an overhang too

Develop software to design and domesticate/optimize the tag aim

Software

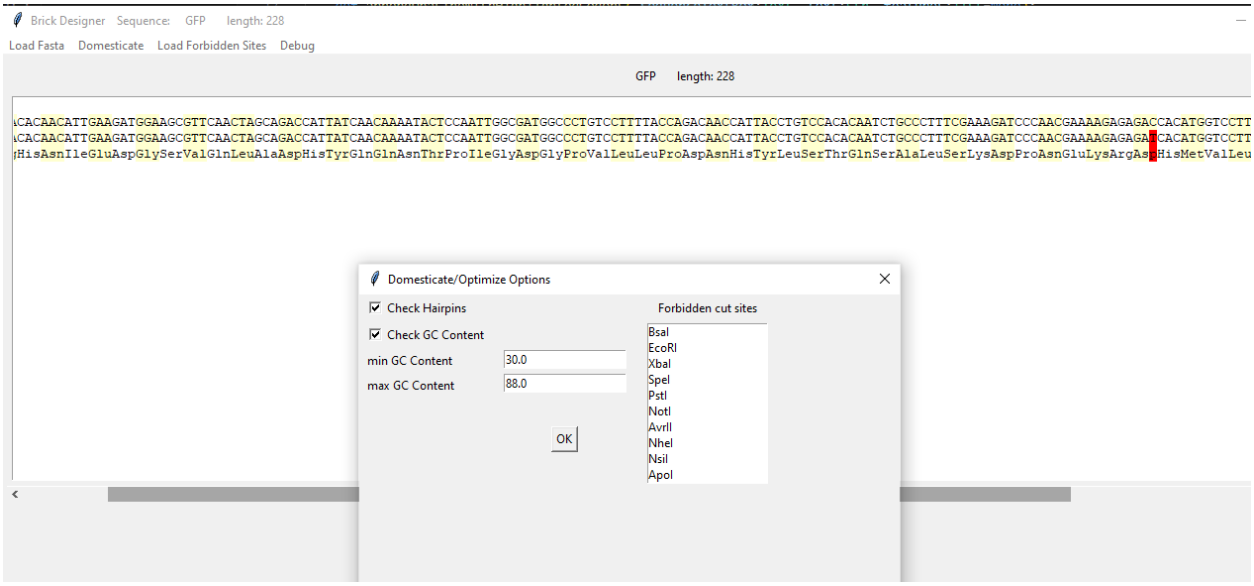
Domestication for Bsa1, Bpi1, and BsmB1 was performed

Checked and (if needed) base replacement was done for hairpins

Checked and (if needed) base replacement was done for GC content and E-Coli

Chi sequence GCTGGTGG was manually checked

Example screenshot from another gene that contains forbidden sequences as the tag does not contain any. You can see a C was replaced with a T but we preserved the **Aspartic acid (Asp).**



N terminal Option 1 (just the affinity Silica tag)

AAAAA ,**GGTCTCAN , CCAT , (domesticated converted MSPHPHPRHHHT to DNA) , NN GTCA, NGAGACC, TTTTT**

GGTCTCAN , CCAT , ATGAGCCCGCATCCGCATCCGCGCCATCATCATACC , NN GTCA, NGAGACC

Note: I could just reused the AT from the prefix and removed the second AT from the start codon of the tag

We choose half randomly some values for NN, in this case AT. We still need to make sure the codon created by NN and G is a viable codon for e-coli so we check that ATG (M)s is a valid codon which it is.

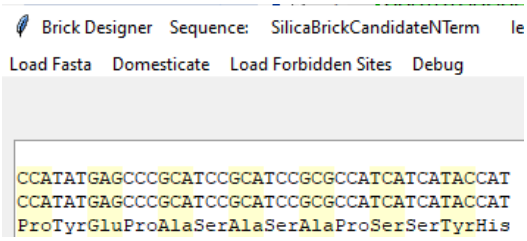
We ALSO remove the first N as the C in the prefix will supply the base for the Bsa1 recognition sequence. We cannot do the same with the last N as the A in suffix because Bsa1 will cut too much and we need the 2 overhangs to be exactly prefix and suffix. So we substitute last N with A

AAAAA, **GGTCTCA , CCAT, ATGAGCCCGCATCCGCATCCGCGCCATCATCATACC, AT GTCA, AGAGACC, TTTTT**

Order for synthesis :

AAAAA**GGTCTCACCAT***ATGAGCCCGCATCCGCATCCGCGCCATCATCATACC***ATGTCAAGAGACCTTTTT**

49%GC



which passes going through the domestication software without changes. Please note we excluded the Bsa1 flanks as they would have been 'domesticated'.

Note: after digestion we will have only ATGAGCCCGCATCCGCATCCGCGCCATCATCATACCATGTCA

C terminal Option 1 (just the affinity Silica tag)

NNNNN **GGTCTCAN, CGGC *NN* , ATGAGCCCGCATCCGCATCCGCGCCATCATCATACC , TAA, GCTT, NGAGACC NNNNN**

we choose half randomly some values for NN, in this case AT:

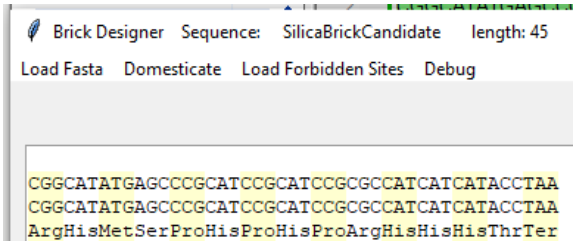
GGTCTCAA, CGGC AT , ATGAGCCCGCATCCGCATCCGCGCCATCATCATACC , TAA GCTT, AGAGACC

so we need to **Order for synthesis** :

GGTCTCAACGGCATATGAGCCCGCATCCGCATCCGCGCCATCATCATACCTAAGCTTAGAGACC

56%GC

which passes going through the domestication software without changes. Please note we excluded the Bsa1 flanks as they would have been 'domesticated'.



Note: initially I intended to perform an alignment of known variants to identify highly conserved regions so we avoid making any changes however as no changes are required we did not need to perform that.

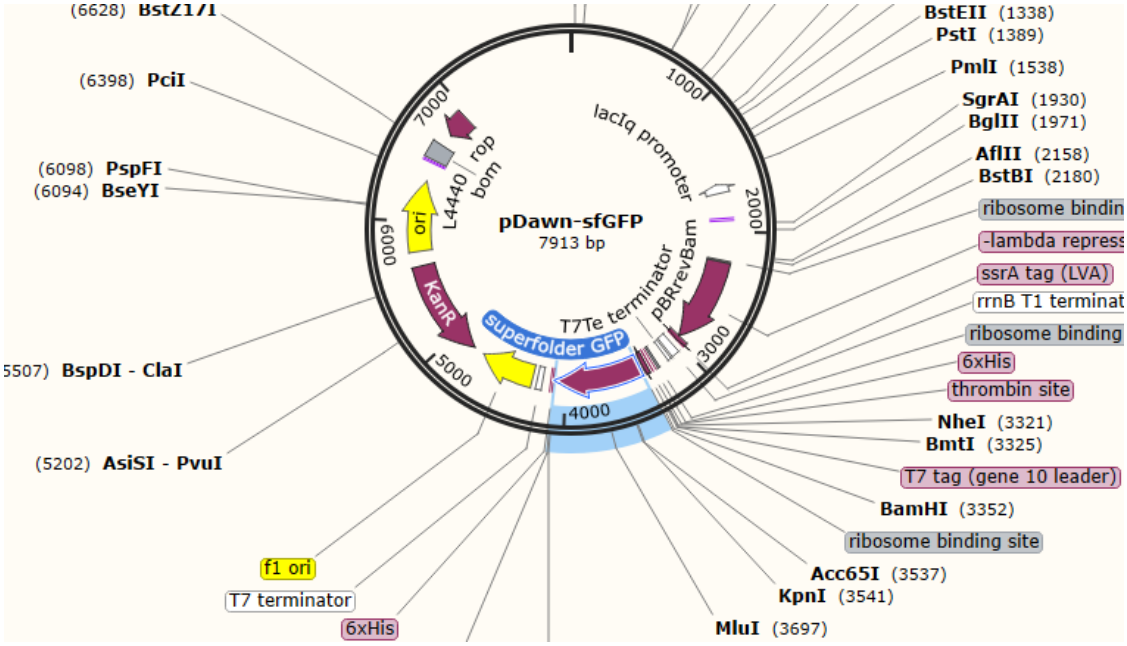
Find an appropriate induction system aim

pDawn, GanBank id is JN579121 and the original one that includes a Fluorescent protein, the author (Ohlendorf) perform tests. (1): From dusk till dawn: one-plasmid systems for light-regulated gene expression. Ohlendorf R et al.
Nucleotide sequences for pDusk and pDawn have been
deposited in the GenBank database under accession numbers JN579120 and JN579121

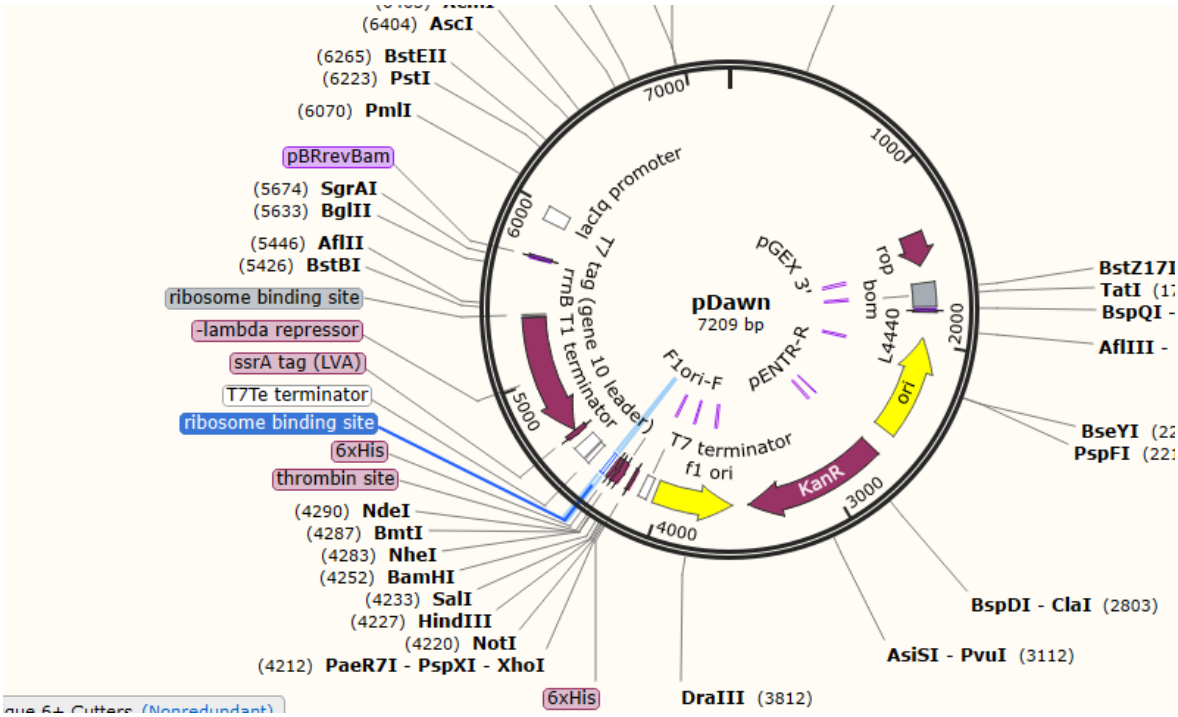
There is one Dawn Module on line 35 in FreeGenes E. coli Protein Expression Toolkit at
https://stanford.freegenes.org/collections/all/products/expression-tookit#description

https://benchling.com/s/seq-k9xuAS64ELSCcH6NMQ8j

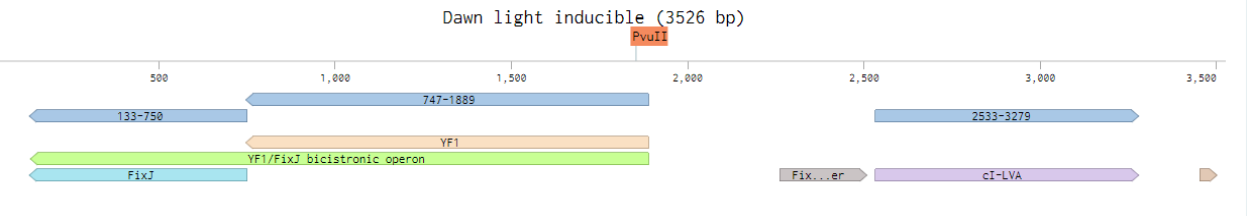
As it's shown in a plasmid we'll use AddGene pDawn for exemplification



AddGene plasmid. with a protein. Our fusion gene must be inserted between RBS and the 6xHis as in this image that contains a GFP insert.



AddGene empty backbone



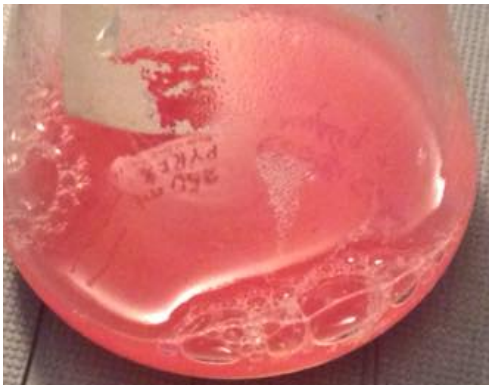
Free GreeGenes pDawn construct

Induction mechanism:

YF1 and FixJ are two genes that are expressed constantly from the LacI promoter(Iq). YF1 is a Histidine Kinase that phosphorylates FixJ in the dark. It contains a Light-Oxygen-Voltage(LOV)
domain that in response to light causes the YF1 protein to change structure and when lit it does not phosphorylate FixJ. We have a bicistronic operon YF1 and FixJ gene with LacI promoter(Iq), RFP with pR promoter, cI a repressor for pR pDawn

NO LIGHT
Dark activates-> YF1 protein that changes structure and become active
activated (new structure) YF1 activates (phosphorylates) FixJ
activated (phosphorylated) FixJ activates (binds) to FixK2 promoter DNA binding site
activated (bound) FixK2 promoter activates (binds) the promoter for the lambda repressor CI and expresses it
activated lambda repressor acts on the pR promoter by repressing (inactivating) it
inactivated (repressed) pR promoter activates (stops production of new mRNA) so no more FP gene

Unit tested the system with the same fluorescent plasmid Ohlendorf used for testing. We obtained this. Guess what color was the fluorescent protein?



An older note specifies “cannot use the EcoRI site in the MCS for cloning as it's in the genomic yet not annotated”

Assembly

After we designed or identified the fusion genes parts we need to assemble the induction system, the gene of interest, the linker BioBrick, affinity tag BioBrick and a backbone plasmid

Order (synthesis) the chosen affinity tag or both affinity tag and linker according to the option of designed sequence.

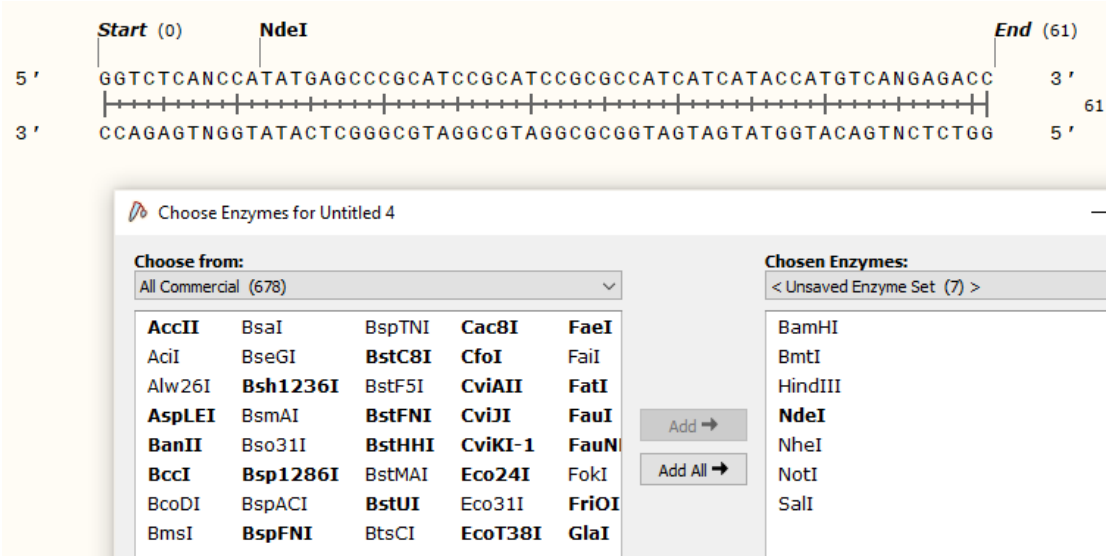
See https://synbio.tsl.ac.uk/uploads/a94d80d88a4d03acbe8b53da74fdcf49.pdf for an excellent cheatsheet on MoCLo assembly

First we do a one pot GoldenGate to connect the linker, Silica affinity tag and a gene of interest available in BioBrick format by using the type 2 Bsa1 enzyme to create the overhangs for hybridization and then use ligase to fill in the nicks. So now the 3 BioBrick fusion gene is created.

Option 1

Insert it in a plasmid that has the pDawn system and promoters terminators etc.

As the backbone plasmid has all the induction we can linearize the plasmid cutting in the MCS by cutting between the RBS and 6xHis tags (assuming AddGene pDawn backbone) with either Bmt1, Nhe1, BamH1, Sal1, HindIII, Not1 or Hho1 and use GA to insert the new fusion gene. Nde1 would have been a choice to but it appears in the silica tag while the others do not:



Option 2

Perform a full part level 2 BioBrick assembly by stringing a BioBrick acceptor plasmid, pDawn Biobrick, our SilicaBindingTag-CleaverLinker-OurGene, a BioBrick Terminator and a BioBrick UTR

Option 3

Just to test that it's working we can skip the standardization and we can order one complete fusion gene sequence with the Bmt1 overhang, silica tag, linker, gene of interest, HindIII overhang then cut the plasmid between 5' Bmt1 and 3' HindIII restriction sites. The overhangs will hybridize and we'll fill in the nicks with ligase.

Testing

Main protocol steps:

Transform the assembled plasmid in BL21 using electroporation, CaCl2 or another chemical transformation like the one used by AddGene.
Plate on Lb-Agar-Kan and incubate at 37 Celsius for 6-12 hours until colonies are formed

Pick one colony and inoculate in LB -Kan liquid media. Place the flask in a incubated shaker until OD600 =.3

Turn on Blue or White light to start induction
Incubate until OD600 =.5
Centrifuge, discard supernatant and resuspend

Lyse an aliquot by sonication
Perform raw purification using centrifuge base separation and enzymes that digest some of the unwanted lysate components like the nucleic acids.

Incubate with silica coated magnetic beads

Use magnet to separate product

Cleave the linker

Use a strong magnet to remove the linker and affinity tag

Ethics, Safety

Goals & Rationale

The goal is to make the production of many expensive reagents faster and cheaper and eliminate cold chain storage and transport needs.

None of the protocols necessary are uncommon to justify additional steps besides normal lab safety protocols. For instance in Canada, research e-coli is considered Risk 1 category that pretty matches BSL1 so at the same level of danger as fridge stored food.

Making research more affordable and accelerate health research, pollution prevention, climate change prevention research etc has an obvious ethics dimension.

Design Requirements

Follow the design to produce the new part and test it following the testing procedure

References

1. “Site-Specific Labeling of Surface Proteins on Living Cells Using Genetically Encoded Peptides that Bind Fluorescent Nanoparticle Probes”, Mark A Rocco 1, Jae-Young Kim, Andrew Burns, Jan Kostecki, Anne Doody, Ulrich Wiesner, Matthew P DeLisa
2. A Golden Gate Modular Cloning Toolbox for Plants;Carola Engler〃 Mark Youles, Ramona Gruetzner,Tim-Martin Ehnert,Stefan Werner,Jonathan D. G. Jones,Nicola J. Patron and Sylvestre Marillonnet
3. From dusk till dawn: one-plasmid systems for light-regulated gene expression. Ohlendorf R, Vidavski RR, Eldar A, Moffat K, Moglich A, J Mol Biol. 2012 Mar 2;416(4):534-42. doi: 10.1016/j.jmb.2012.01.001. Epub 2012 Jan 8
4. Silica-Precipitating Peptides Isolated from a Combinatorial Phage Display Peptide Library Rajesh R. Naik,a Lawrence L. Brott,a Stephen J. Clarson,b and Morley O. Stonea
5. Joint Universal Modular Plasmids (JUMP): A flexible and comprehensive platform for synthetic biology, https://www.biorxiv.org/content/10.1101/799585v1.full

View ORCID ProfileMarcos Valenzuela-Ortega, View ORCID ProfileChristopher French

1. **A cleavable silica-binding affinity tag for rapid and inexpensive protein purification** Brandon L Coyle, François Baneyx
2. Removal of Affinity Tags with TEV Protease -Sreejith Raran-Kurussi, Scott Cherry, Di Zhang, and David S. Waugh