**Instructivo**

Por: Gerardo Cendejas Mendoza

Este es un instructivo para realizar la búsqueda de enzimas candidatas para los ensayos enzimáticos que se realizarán en conjunto con Ginkgo Bioworks. Las enzimas seleccionadas para este paso fueron las enzimas 5, 6, 9, 10 y 11.

| Enzima | locus ID | EC number | InterPro |
| --- | --- | --- | --- |
| [5](https://drive.google.com/drive/folders/134K3-a-zAcIUCO9z0QJ3FyfvJBcS-WVf?usp=drive_link) | Pn7.1626 | EC:1.14.19.- | PF00067 |
| [6](https://drive.google.com/drive/folders/1SFH57YKP92hESHVYEfOyv0kGpkM9u4gM?usp=drive_link) | Pn16.1198 | EC:6.2.1.12 | PF00501  PF13193  PF13456  PF00141  PF03398 |
| [9](https://drive.google.com/drive/folders/13SQ-e5PdmnAuloWKqFlQhKoSspS3n-2U?usp=drive_link) | Pn2.2377 |  | PF03936  PF01397  PF00574  PF13976  PF14223  PF00141 |
| [10](https://drive.google.com/drive/folders/10gENPjMJHsj0ZjO-DY4EpW2Bbhe0N-Xc?usp=drive_link) | N/A | EC:1.5.1.21  EC:1.4.1.17 |  |
| [11](https://drive.google.com/drive/folders/1C8USl11jaZP9kz1p03oGx5hrrVNZU0-l?usp=drive_link) | Pn6.2477 | EC:2.3.1.196  EC:2.3.1.232 | PF02458 |

Dentro de la carpeta en que se encuentra este documento hay también carpetas para cada una de las enzimas seleccionadas. En estas se encuentra ya la secuencia de DNA y de aminoácidos para las enzimas seleccionadas, el archivo (enzima).fas contiene la secuencia de nucleótidos optimizada para *Saccharomyces cerevisiae*, mientras que el archivo (enzima)\_aa.fas contiene la secuencia de aminoácidos de la proteína.

Secciones:

* Encontrar enzimas con la actividad deseada usando bases de datos
* Encontrar enzimas por homología
* Hacer una librería

**Bases de datos**

1. Se realiza la búsqueda del EC number en nuestra anotación del genoma ([aquí](https://github.com/GerardoCMM/Synthetic-Biobots-IGEM-DL-2022/blob/main/Anot/Genome/merged.txt)), en la columna ko description, está entre paréntesis uno o más EC number que fueron asignados a dicha enzima, buscar (Ctrl+F) el ID de la enzima.

En caso de no encontrar EC number en el documento, contactarse directamente con Gerardo para determinar el procedimiento a seguir.

1. Realizar la búsqueda en [Uniprot](https://www.uniprot.org/) de los EC number encontrados (Una búsqueda por cada EC), colocar en el buscador en el formato (ec:x.x.x.x), descargar en formato TSV y subir a la carpeta de la enzima trabajada (guardar con el nombre (enzima)\_Uniprot\_EC:x.x.x.x.tsv).
2. En la base de datos [Brenda](https://www.brenda-enzymes.org/) buscar el EC number en formato x.x.x.x, en los resultados abrir el HIT para EC number. En la parte izquierda de la pantalla seleccionar “enzyme structure” y hacer click en “show complete table” y “print all entries” posteriormente. Guardar como (enzima)\_Brenda\_EC:x.x.x.x.pdf y subir a la carpeta correspondiente.
3. Realizar la búsqueda en [InterPro](https://www.ebi.ac.uk/interpro/search/text/), para esto se requiere saber la clase de enzima según InterPro, esta debe buscarse en este archivo ([aquí](https://github.com/GerardoCMM/Synthetic-Biobots-IGEM-DL-2022/blob/main/RNAseq/Genes_Anot.ods), descargar el archivo para visualizarlo en libre office o excel). Buscar el ID de la enzima trabajada y en la segunda columna de “Pfam Domains” entre paréntesis se encuentra la familia de InterPro de la enzima. Buscar esa familia en InterPro. Abrir el resultado de la familia, ir al apartado proteins, seleccionar reviewed y exportar en formato tsv, guardar como (enzima)\_InterPro\_(familia).tsv.

**Búsqueda por homología**

1. Realizar un [blastp](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) contra la base de datos del NCBI. Usar como referencia la secuencia de aminoácidos que se encuentra en la carpeta de la respectiva enzima. Dejar los parámetros por defecto y correr el blastp. Descargar los primeros 100 hits, la “Descriptions table” en formato csv y guardar como (enzima)\_blastp\_ncbi.csv.
2. Realizar un blast contra la base de datos UniProt ([aquí](https://www.uniprot.org/blast)). Pegar la secuencia de aminoácidos de la enzima trabajada en formato fasta. Dejar los parámetros por defecto y correr el programa. Abrir los resultados y descargar el overview en formato tsv y guardar con el nombre (enzima)\_blast\_uniprot.csv
3. Encontrar más enzimas por homología utilizando [EVcouplings](https://v2.evcouplings.org/), pegar la secuencia de aminoácidos en formato fasta, escribir un nombre de trabajo que incluya el número de enzima estudiada y un correo electrónico para recibir los resultados.

Progreso…

| Enzima | Búsqueda por homología | Bases de datos |
| --- | --- | --- |
| [5](https://drive.google.com/drive/folders/134K3-a-zAcIUCO9z0QJ3FyfvJBcS-WVf?usp=drive_link) | * ~~NCBI~~ * ~~UniProt~~ * EVcouplings | * ~~UniProt~~ * ~~Brenda~~ * ~~InterPro~~ |
| [6](https://drive.google.com/drive/folders/1SFH57YKP92hESHVYEfOyv0kGpkM9u4gM?usp=drive_link) | * ~~NCBI~~ * ~~UniProt~~ * EVcouplings | * ~~UniProt~~ * ~~Brenda~~ * ~~InterPro 1~~ * InterPro 2 |
| [9](https://drive.google.com/drive/folders/13SQ-e5PdmnAuloWKqFlQhKoSspS3n-2U?usp=drive_link) | * ~~NCBI~~ * ~~UniProt~~ * EVcouplings | * UniProt * UniProt * ~~Brenda~~ * InterPro * InterPro |
| [10](https://drive.google.com/drive/folders/10gENPjMJHsj0ZjO-DY4EpW2Bbhe0N-Xc?usp=drive_link) | * ~~NCBI~~ * ~~UniProt~~ * ~~EVcouplings~~ | * ~~UniProt 1~~ * ~~UniProt 2~~ * ~~Brenda~~ * InterPro 1 * InterPro 2 |
| [11](https://drive.google.com/drive/folders/1C8USl11jaZP9kz1p03oGx5hrrVNZU0-l?usp=drive_link) | * ~~NCBI~~ * ~~UniProt~~ * ~~EVcouplings~~ | * ~~UniProt~~ * ~~Brenda~~ * ~~InterPro~~ |