

UNIVERSIDAD DE GUADALAJARA

---

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍA  
DIVISIÓN DE ELECTRÓNICA Y COMPUTACIÓN



ASOCIACIÓN POR APRENDIZAJE PROFUNDO DE K-MEROS GENÓMICOS CON ETIQUETAS  
ONTOLÓGICAS

PROTOCOLO DE POSGRADO

MAESTRÍA EN CIENCIAS EN BIOINGENIERÍA Y CÓMPUTO INTELIGENTE

PRESENTA

DIANA LAURA GONZÁLEZ RÍOS

DIRECTOR DE TESIS

DR. J. ALEJANDRO MORALES VALENCIA

CODIRECTORA

DRA. ADRIANA PATRICIA MENDIZABAL RUIZ

GUADALAJARA, JALISCO. 02 DE MARZO DEL 2022

## **Abstract**

### **Abstract:**

Currently, more information about DNA has been discovered that the traditional study structure does not allow us to take full advantage of it, so efforts should be directed at a new approach to process the information that we already have and take advantage of technological resources to innovate in this area. A good starting point would be to analyze the functional structure of genes with a new eye, taking advantage of the existing ontological classification and software tools that allow the extraction of essential functional fragments of these chains with applications in new technologies such as tools for gene editing.

### **Resumen:**

Actualmente se ha descubierto más información sobre el DNA que la estructura de estudio tradicional no nos permite aprovechar completamente, por lo que se deben dirigir esfuerzos en un nuevo enfoque para procesar la información que ya se tiene y aprovechar los recursos tecnológicos para innovar en esta área. Un buen punto de inicio sería analizando con nuevo ojo la estructura funcional de los genes, aprovechando la clasificación ontológica existente y las herramientas de software que permitan extraer los fragmentos funcionales esenciales de estas cadenas con aplicaciones en nuevas tecnologías tales como herramientas para la edición génica.

## **Descripción:**

Los genes contienen información codificante de organismos vivos y sus productos contienen todos los tipos de funciones en la célula[1,2], lo que permite asignar un conjunto de etiquetas que describen dichas funciones biológicas llamadas ontologías[3].

Típicamente, las ontologías se asignan a toda la secuencia del gen pero existen dos problemas principales con esta aproximación:

(1) Aunque las etiquetas ontológicas se asocian al gen completo, con frecuencia en bioinformática las funciones se asocian a los dominios. Los dominios son subsecuencias homólogas resultado del alineamiento de los genes[4]. Estos dominios explican de manera parcial el tipo de actividad que tendrán estos genes[5]. Al final se puede considerar que la suma y estructura de los dominios dará una pista sobre la función del gen en la célula[6].

(2) La comparación entre secuencias homólogas ha demostrado que el porcentaje de similitud entre secuencias que tienen funciones o actividades semejantes es muy variable y con frecuencia no es posible identificar patrones estadísticos de cambio entre dichas secuencias. Y se ha concluido que el alineamiento no es la herramienta para la identificación de dominios o para la atribución de funciones de los genes.

Hoy en día existen técnicas que permiten encontrar más y mejores asociaciones a funciones dentro de secuencias de genes que no se basan en el alineamiento y búsqueda de dominios como las técnicas de Machine Learning[7,8]. En este trabajo propongo desarrollar una herramienta de deep learning que permita atribuir etiquetas ontológicas a segmentos de secuencias génicas.

## **Trabajos relacionados:**

Cuando se habla de Deep Learning aplicado a genómica se pueden ver distintas aplicaciones de acuerdo a las cualidades de los modelos y a la información que se quiere obtener de estos experimentos [9].

En ejercicios de este tipo se observan trabajos que utilizan la información ya disponible sobre dominios proteicos, secciones muy definidas y con datos que son respaldados por experimentos y asociación estadística.

Usando como referencia el trabajo de Borrayo et al (2020) [10] como referencia para analizar los genes como una secuencia fragmentable donde la información es extraíble extractos representativos de tamaño 13, usando tales fragmentos como la unidad mínima de secuencia o “palabra” dentro de un gen bacteriano dado.

Dado que las “palabras” génicas se van a asociar a etiquetas ontológicas de manera que encontremos subconjuntos o grupos de palabras definibles como “palabras clave” para extraer el mensaje que transmite un gen y que se ve reflejado en la o las funciones asociadas a éste en su organismo de origen.

Trabajos como DeepBind [11], DeepSEA [12] y Basset [13], y otras publicaciones como Yin et al [14], analizan secuencias de DNA utilizando fragmentos o extractos de las secuencias, utilizando redes neuronales convolucionales (CNN por sus siglas en inglés) para encontrar patrones en estas secuencias, exponiendo la utilidad de estos algoritmos en la predicción de funciones en áreas biológicas.

Existen otros algoritmos que pueden complementar a las CNN, como es el caso de los módulos de concepto-transformador (CT por sus siglas en inglés) que son capaces de mejorar la toma de decisión de una red neuronal extrayendo patrones internos en las características extraídas del vector, añadiendo conceptos o definiciones más características a las clases que trabaje el learner [15], demostrándose bastante confiables y comprobables por medio de otros algoritmos de machine learning [16].

El problema descrito antes también puede abordarse desde una perspectiva de análisis de texto, aprovechando algoritmos que ya han trabajado en la definición de palabras por medio de ontologías que describen tanto las palabras en sí, como sus relaciones con otras, existiendo también la posibilidad de utilizar redes neuronales enfocadas al análisis y descripción de texto por ontologías[17], siendo este mismo tipo de algoritmos también aplicables al análisis de DNA como el caso de deepTarget[18], deepMiRGene[19] que usan redes neuronales recurrentes (RNN por sus siglas en inglés) o el caso de Jiang et al (2016) [20] y Rachmatia et al (2017)[21] que usan redes de creencia profunda (DBN por sus siglas en inglés) para analizar secuencias genómicas del mismo que el análisis de texto.

## **Hipótesis:**

Hay fragmentos de genes significativamente asociados a sus etiquetas ontológicas.

## **Metas y objetivos:**

Meta 1: Relacionar las etiquetas ontológicas con subsecuencias de genes.

Objetivo 1: Obtener un modelo que asocie etiquetas ontológicas con k-meros.

Objetivo 2: Correlacionar estos k-meros con los dominios ya existentes en los genes.

### Metodología propuesta y cronograma:

AÑO	2022																												
MES	Enero				Febrero				Marzo					Abril				Mayo				Junio							
SEMANA	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	3	4	1	2	3	4	5			
Minar las secuencias de genes y sus ontologías.																													
Preprocesar los datos.																													
Elegir las arquitecturas y ajustarlas.																													
Hacer el entrenamiento, validación y testing.																													
Separar los kmeros de acuerdo a su orientación ontológica.																													
Ensamblar los kmeros.																													
Alinearlos contra dominios conocidos.																													
Escritura de tesis																													
Envío de artículo																													
Posteo de método																													
Defensa de tesis																													
MES	Julio				Agosto					Septiembre				Octubre				Noviembre					Diciembre						
SEMANA	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4			
Minar las secuencias de genes y sus ontologías.																													
Preprocesar los datos.																													
Elegir las arquitecturas y ajustarlas.																													
Hacer el entrenamiento, validación y testing.																													
Separar los kmeros de acuerdo a su orientación ontológica.																													
Ensamblar los kmeros.																													
Alinearlos contra dominios conocidos.																													
Escritura de tesis																													
Envío de artículo																													
Posteo de método																													
Defensa de tesis																													

Figura 1. Cronograma de metodología propuesta.

## Metodología

### Base de datos

Se ha seleccionado para trabajar una base de datos del consorcio de Gene Ontology[1], específicamente la perteneciente a la Enciclopedia de metabolismo de *E. coli*[2] debido a que la información contenida en este archivo fue inferida por medios experimentales[3] con 55232 anotaciones hasta la fecha, de cuyo contenido se tomaron los siguientes datos:

- DB
- DB Object ID
- GO ID

Se procesó la información de este archivo para unir todas las anotaciones ontológicas por gen, obteniendo un total de 3920 datos, y utilizando el ID correspondiente a las bases de datos de Uniprot se accedió a la información de ésta por medio de un algoritmo en python. Las secuencias obtenidas eran de las proteínas producto de los genes de *E. coli*, y se procesó cada secuencia por medio de una retrotraducción.

**Tabla 1:** Codones de aminoácidos en su equivalente a secuencia de DNA. Se sustituyó el Uracilo por Timina en los casos que aplicaba.

Aminoácido	Abreviación	Codones
Alanina	A	GCT, GCC, GCA, GCG
Cisteína	C	TGT, TGC
Ácido Aspártico	D	GAC, GAT
Ácido Glutámico	E	GAA, GAG
Fenilalanina	F	TTT, TTC
Glicina	G	GGT, GGC, GGA, GGG
Histidina	H	CAT, CAC
Isoleucina	I	ATT, ATC, ATA
Lisina	K	AAA, AAG
Leucina	L	TTA, TTG, CTT, CTC, CTA, CTG
Metionina	M	ATG
Asparagina	N	AAC, AAT
Prolina	P	CCT, CCC, CCA, CCG
Glutamina	Q	CAA, CAG
Arginina	R	AGA, AGG, CGA, CGC, CGG, CGT
Serina	S	TCT, TCC, TCA, TCG, AGT, AGC
Treonina	T	ACT, ACC, ACA, ACC
Selenocisteína	U	TGA
Valina	V	GTT, GTC, GTA, GTG
Triptófano	W	TGG
Tirosina	Y	TAT, TAC
Otro glutámico	Z	GAA, GAG

Para la traducción inversa se generó un diccionario de equivalencias de aminoácidos con sus equivalencias en codones de DNA como se ve en la **Tabla 1**, de acuerdo al código genético estándar[4], con la excepción de el añadido de un codón para el aminoácido Selenocisteína, que normalmente es interpretado como un codón de paro.

Se fragmentaron las posibles secuencias de DNA de cada proteína de la base de datos con una superposición de 1 nucleótido para obtener cada substring de tamaño 13 que corresponde a los kmeros de la secuencia de DNA. Con esta información se encontraron un total de 66129041 kmeros únicos, representando un 98.53% de los kmeros totales posibles por combinación para el tamaño del kmero.

La base de datos se preparó para ser utilizada, de manera que tenemos un conjunto de kmeros asociados a conjuntos de ontologías acordes a las proteínas donde apreciaron

dichos kmeros, revolviendo el orden en el que aparecen dichos datos pero manteniendo la asociacion para alimentar el modelo. Los datos se alimentarán con la proporción:

- 89% Training
- 9% Testing
- 2% Validation

## Estadística

Analizando la distribución de frecuencias de las etiquetas ontológicas obtenemos un total de 4206 etiquetas en esta base de datos de las 43588 etiquetas actualizadas y activas existentes en el catálogo de Gene ontology[5], obteniendo el histograma que se muestra a continuación.



Figura 2. Histograma de frecuencia de ontologías de *E. coli*.

Dentro de las estadísticas de las ontologías presentes en los genes de esta base de datos se tiene lo siguiente: En los conteos de ontologías por gen el mínimo de ontologías por gen es 1, dado en 199 casos, mientras que el máximo de etiquetas por gen es 34, con un valor promedio de 8.5 etiquetas por gen. Además Analizando las ontologías observamos que el mínimo de aparición es 1 en 1996 casos, casi el 50% de las ontologías

totales de este grupo, y el máximo de aparición por ontología es 1320 con un promedio de aparición de 7.95 veces.

## Modelo

Analizando los trabajos relacionados y la naturaleza de la información se ha optado por dos arquitecturas de redes neuronales para procesar la información, una CNN y una Bidirectional long-short term memory RNN (BLSTM RNN por sus siglas en inglés). Ambas redes serán entrenadas y comparadas en rendimiento y resultados.

Cada kmero es procesado para generar un vector One Hot Encode (OHE), generando un vector de 4\*13 para la CNN y un vector 1\*52 para la BLSTM RNN, así como se hizo el mismo procedimiento para generar el vector de ontologías correspondiente a la salida de cada red neuronal, generando para cada proteína un vector OHE de ontologías que será usado cada kmero que entre a la red de acuerdo a la proteína de procedencia.

Debido a la naturaleza del vector de salida al crearse un vector para los genes de la base de datos se obtienen vectores de máximo 34 unos contra 43554 ceros, haciendo que si el modelo clasifica como cero una etiqueta donde debiera haber un uno el error es mínimo, por ende se aprovechará la métrica de área bajo la curva ROC, que analiza la clasificación de la información entre verdaderos positivos y verdaderos negativos (TP y TN respectivamente)[6][7], mostrando la capacidad del modelo de acertar en la clasificación de pertenencia a cada clase (etiqueta ontológica) posible.

Se espera identificar los kmeros

## CNN

Para la arquitectura de esta red neuronal se ha usado un modelo para clasificar números escritos a mano[8], modificando los metaparámetros para ajustar el número de neuronas según el número de ontologías posibles donde pudiera relacionarse cada kmero ingresado. La arquitectura resultante es la siguiente:

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 4, 13, 100)	2600
max_pooling2d (MaxPooling2D)	(None, 2, 6, 100)	0
dropout (Dropout)	(None, 2, 6, 100)	0
conv2d_1 (Conv2D)	(None, 2, 6, 1000)	901000
max_pooling2d_1 (MaxPooling2D)	(None, 1, 3, 1000)	0
dropout_1 (Dropout)	(None, 1, 3, 1000)	0
flatten (Flatten)	(None, 3000)	0
dense (Dense)	(None, 20000)	60020000
dropout_2 (Dropout)	(None, 20000)	0
dense_1 (Dense)	(None, 43558)	871203558

=====  
Total params: 932,127,158

Trainable params: 932,127,158

Non trainable params: 0

Figura 3. Arquitectura de red neuronal convolucional.

## BIBLIOGRAFÍA METODOLOGÍA

1. Gene Ontology Consortium. (s. f.). Download Annotations for 2022–06-15 release. Gene Ontology Downloads. Recuperado 5 de julio de 2022, de <http://current.geneontology.org/products/pages/downloads.html>
2. Azweifel. (2017, 20 julio). Welcome to EcoliWiki. EcoliWiki. Recuperado 5 de julio de 2022, de [https://ecoliwiki.org/colipedia/index.php/Welcome\\_to\\_EcoliWiki](https://ecoliwiki.org/colipedia/index.php/Welcome_to_EcoliWiki)
3. Gene Ontology Consortium. (2018, 23 febrero). Guide to GO evidence codes. Gene Ontology Resource. Recuperado 5 de julio de 2022, de <http://geneontology.org/docs/guide-go-evidence-codes/>
4. Elzanowski, A. A., Ostell, J., & National Center for Biotechnology Information. (2019, 7 enero). The Genetic Codes. NCBI - The Genetic Codes. Recuperado 9 de agosto de 2022, de <https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>
5. Gene Ontology Consortium. (s. f.-a). AmiGO 2: Search. AmiGO 2. Recuperado 5 de julio de 2022, de <http://amigo.geneontology.org/amigo/search/ontology>
6. Gonzalez, L. (2020, 21 agosto). Curvas ROC y Área bajo la curva (AUC). Aprende IA. Recuperado 29 de agosto de 2022, de <https://aprendeia.com/curvas-roc-y-area-bajo-la-curva-auc-machine-learning/>
7. Gonzalez, L. (2020b, agosto 21). Matriz de Confusión. Aprende IA. Recuperado 29 de agosto de 2022, de <https://aprendeia.com/matriz-de-confusion-machine-learning/>
8. Jang, A. (2020, 9 junio). TensorFlow: MNIST CNN Tutorial. Kaggle. Recuperado 29 de agosto de 2022, de <https://www.kaggle.com/code/amyjang/tensorflow-mnist-cnn-tutorial/notebook>
- 9.

## Resultados esperados e impacto:

Esperamos concluir con un método que nos dé la oportunidad de explorar las secuencias génicas y las funciones en las que éstas se involucran. Con este método esperamos tener un impacto sobre toda la comunidad científica que trabaja en ciencias ómicas, así como en el campo de la edición genómica. Finalmente los resultados de este trabajo se publicarán en una revista internacional de alto prestigio.

## Discusion:

Existen dos herramientas dentro de los servicios del EMBL[1,2] relacionados a proteínas que devuelven la traducción inversa de las proteínas. La primera herramienta es EMBOSS backtranambig traduce la secuencia de aminoácidos a una única secuencia ambigua de DNA[3], utilizando el Fasta de la proteína y la selección de una tabla de codones de acuerdo a los códigos genéticos existentes[4-Met4], con una salida de una



secuencia de DNA basada en la raducción inversa más favorable de acuerdo a sus registros[5]. La segunda herramienta es EMBOSS Backtranseq[6]

### **Bibliografía Discusion:**

1. European Bioinformatics Institute. (s. f.). Data resources and tools. EMBL's European Bioinformatics Institute. Recuperado 16 de agosto de 2022, de <https://www.ebi.ac.uk/services/data-resources-and-tools?category=proteins>
2. Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research, 50(W1), W276-W279. <https://doi.org/10.1093/nar/gkac240>
3. EMBL-EBI. (s. f.). EMBOSS Backtranambig. EMBOSS Backtranambig < Sequence Translation Sites < EMBL-EBI. Recuperado 16 de agosto de 2022, de [https://www.ebi.ac.uk/Tools/st/emboss\\_backtranambig/](https://www.ebi.ac.uk/Tools/st/emboss_backtranambig/)
4. Elzanowski, A. A., Ostell, J., & National Center for Biotechnology Information. (2019, 7 enero). The Genetic Codes. NCBI - The Genetic Codes. Recuperado 9 de agosto de 2022, de <https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>
5. Bleasby, A. (s. f.). EMBOSS: backtranambig. EMBOSS backtranambig. Recuperado 16 de agosto de 2022, de <http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/backtranambig.html>
6. EMBL-EBI. (s. f.-b). EMBOSS Backtranseq. EMBOSS Backtranseq < Sequence Translation Sites < EMBL-EBI. Recuperado 16 de agosto de 2022, de [https://www.ebi.ac.uk/Tools/st/emboss\\_backtranseq/](https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/)

### **Factibilidad y seguimiento:**

Este trabajo es continuidad del trabajo de tesis doctoral del Dr. Isaías May Cache, realizándose de manera individual y sin ningún tipo de financiamiento. Para su realización se utilizarán los recursos:

- Python 3.8
- Equipo de cómputo en laboratorio
- Google colab
- Gene Ontology - Biopython

### **Referencias Descripción y Trabajos relacionados**

1. Stotz, K., Griffiths, P. E., & Knight, R. (2004). How biologists conceptualize genes: an empirical study. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 35(4), 647–673. <https://doi.org/10.1016/j.shpsc.2004.09.005>
2. Zhang, M., Perelson, A. S., & Tung, C. (2011). RNA Structural Motifs. eLS. Published. <https://doi.org/10.1002/9780470015902.a0003132.pub2>

3. Dessimoz, C., & ŠKunca, N. (Eds.). (2017). The Gene Ontology Handbook. Methods in Molecular Biology. Published. <https://doi.org/10.1007/978-1-4939-3743-1>
4. COMAV. (s. f.). Alineamiento de secuencias. Bioinformatics at COMAV 0.1 documentation. Recuperado 10 de noviembre de 2021, de [https://bioinf.comav.upv.es/courses/intro\\_bioinf/alineamientos.html](https://bioinf.comav.upv.es/courses/intro_bioinf/alineamientos.html)
5. GuateQuímica. (s. f.). Dominios y Motivos. Recuperado 10 de noviembre de 2021, de [http://www.guatequimica.com/tutoriales/proteinas/Dominios\\_y\\_Motivos.htm](http://www.guatequimica.com/tutoriales/proteinas/Dominios_y_Motivos.htm)
6. Bateman, A., Coggill, P., & Finn, R. D. (2010). DUFs: families in search of function. Acta Crystallographica Section F Structural Biology and Crystallization Communications, 66(10), 1148–1152. <https://doi.org/10.1107/s1744309110001685>
7. Morales, J. A., Saldaña, R., Santana-Castolo, M. H., Torres-Cerna, C. E., Borrayo, E., Mendizabal-Ruiz, A. P., Vélez-Pérez, H. A., & Mendizabal-Ruiz, G. (2020). Deep Learning for the Classification of Genomic Signals. Mathematical Problems in Engineering, 2020, 1–9. <https://doi.org/10.1155/2020/7698590>
8. Petrosino, A., Loia, V., & Pedrycz, W. (2017). Fuzzy Logic and Soft Computing Applications. Springer Publishing. <https://doi.org/10.1007/978-3-319-52962-2>
9. Eraslan, G., Avsec, I., Gagneur, J., & Theis, F. J. (2019a). Deep learning: new computational modelling techniques for genomics. Nature Reviews Genetics, 20(7), 389–403. <https://doi.org/10.1038/s41576-019-0122-6>
10. Borrayo, E., May-Canche, I., Paredes, O., Morales, J. A., Romo-Vázquez, R., & Vélez-Pérez, H. (2020). Whole-Genome k-mer Topic Modeling Associates Bacterial Families. Genes, 11(2), 197. <https://doi.org/10.3390/genes11020197>
11. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology, 33(8), 831–838. <https://doi.org/10.1038/nbt.3300>
12. Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. Nature Methods, 12(10), 931–934. <https://doi.org/10.1038/nmeth.3547>
13. Kelley, D. R., Snoek, J., & Rinn, J. L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Research, 26(7), 990–999. <https://doi.org/10.1101/gr.200535.115>
14. Yin, B., Balvert, M., Zambrano, D., Schönhuth, A., & Bohte, S. (2018). An image representation based convolutional network for DNA classification. arxiv. <https://doi.org/10.48550/arXiv.1806.04931>
15. Khan, S., Naseer, M., Hayat, M., Waqas Zamir, S., Shahbaz Khan, F., & Shah, M. (2022). Transformers in Vision: A Survey. arxiv. <https://doi.org/10.48550/arXiv.2101.01169>
16. Rigotti, M., Mikšovic, C., Giurgiu, I., Gschwind, T., & Scotton, P. (2022). ATTENTION-BASED INTERPRETABILITY WITH CONCEPT TRANSFORMERS. ICLR. <https://openreview.net/pdf?id=kAa9eDS0RdO>
17. Al-Aswadi, F. N., Chan, H. Y., & Gan, K. H. (2019). Automatic ontology construction from text: a review from shallow to deep learning trend. Artificial Intelligence Review, 53(6), 3901–3928. <https://doi.org/10.1007/s10462-019-09782-9>

18. Lee, B., Baek, J., Park, S., & Yoon, S. (2016). deepTarget: End-to-end Learning Framework for microRNA Target Prediction using Deep Recurrent Neural Networks. arxiv. <https://doi.org/10.48550/arXiv.1603.09123>
19. Park, S., Min, S., Choi, H., & Yoon, S. (2016). deepMiRGene: Deep Neural Network based Precursor microRNA Prediction. arxiv. <https://doi.org/10.48550/arXiv.1605.00017>
20. Jiang, M., Liang, Y., Feng, X., Fan, X., Pei, Z., Xue, Y., & Guan, R. (2016). Text classification based on deep belief network and softmax regression. *Neural Computing and Applications*, 29(1), 61–70. <https://doi.org/10.1007/s00521-016-2401-x>
21. Rachmatia, H., Kusuma, W. A., & Hasibuan, L. S. (2017). Prediction of maize phenotype based on whole-genome single nucleotide polymorphisms using deep belief networks. *Journal of Physics: Conference Series*, 835, 012003. <https://doi.org/10.1088/1742-6596/835/1/012003>

My goal is to be a scientist and I'm truly focused on that. To successfully complete that goal I need to improve myself frequently, including the abilities that may be considered enough already, and performing my English levels not only to be capable of reading or studying, but to communicate to the world the projects that I'm doing, or do appropriate networking for creating research collaboration is the main reason for why I'm applying here.