

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California).

La investigación fue financiada por el SECIHTI (Secretaría de Ciencia, Humanidades, Tecnología e Innovación).

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México). El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo o titular de los Derechos de Autor.

CICESE © año de defensa, Todos los Derechos Reservados, CICESE

Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California



**Maestría en Ciencias
en Nombre del Posgrado con orientación en ...**

Diseño de proteínas mediante algoritmos de optimización multiobjetivo

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

David Gerardo Murillo Benítez

Ensenada, Baja California, México
año de defensa

Tesis defendida por

David Gerardo Murillo Benítez

y aprobada por el siguiente Comité

Dr./Dra. Xxxx Xxxx Xxxx

Director de tesis

Dr./Dra. Xxxx Xxxx Xxxx

Dr./Dra. Xxxx Xxxx Xxxx

Dr./Dra. Xxxx Xxxx Xxxx



Dr./Dra. Xxxx Xxxx Xxxx

Coordinador del Posgrado en Nombre del Posgrado

Dra. Ana Denise Re Araujo

Directora de Estudios de Posgrado

Resumen de la tesis que presenta David Gerardo Murillo Benítez como requisito parcial para la obtención del grado de Maestro en Ciencias en Nombre del Posgrado con orientación en

Diseño de proteínas mediante algoritmos de optimización multiobjetivo

Resumen aprobado por:

Dr./Dra. Xxxx Xxxx Xxxx

Director de tesis

El resumen, es uno de los elementos más importantes de la tesis, en donde se sintetiza: el objetivo de la investigación, los métodos, los resultados obtenidos y las conclusiones más importantes. El resumen se escribe una vez finalizada la redacción de la tesis. Debe ser completo en cuanto al contenido del trabajo, comprensible, sencillo y preciso. No debe contener citas bibliográficas, figuras o tablas. Respecto al formato, debe escribirse con letra Calibri 11, a renglón sencillo, en un solo párrafo, no superar las 350 palabras en el caso de las tesis de maestría y las 500 en las de doctorado. El número de palabras también aplica para el resumen en inglés.

Palabras clave: pc1, pc2, ..., pc5

Abstract of the thesis presented by David Gerardo Murillo Benítez as a partial requirement to obtain the Master of Science degree in Name of the Degree with orientation in

Title of thesis

Abstract approved by:

PhD Xxxx Xxxx Xxxx
Thesis Director

El abstract es la versión en inglés del resumen en español, por lo que debe proveer la misma información, sin ser una traducción literal. Se aconseja que sea revisado por un investigador o académico que conozca del tema y que domine el idioma inglés.

Keywords: k1, k2, ..., k5

Dedicatoria

Se puede incluir una dedicatoria como homenaje y agradecimiento a quienes hayan influido en el éxito de la investigación. Únicamente texto.

Agradecimientos

A mi madre y hermano, por apoyarme incondicionalmente en todo momento. A mis amigos, por su comprensión y apoyo durante este proceso.

A mis asesores de tesis, Dr. Carlos Brizuela y Dr. Jesus Guillermo Falcon

En la sección de agradecimientos se debe reconocer la ayuda de personas que aportaron significativamente al desarrollo profesional de la investigación. Primero, se incluyen los nombres de las instituciones que financiaron la investigación, especialmente al CICESE y SECIHTI. Únicamente texto.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	vii
Lista de tablas	viii
 Capítulo 1. Introducción	
1.1. Antecedentes	1
1.2. Justificación (opcional)	2
1.3. Hipótesis (opcional)	2
1.4. Objetivos	2
1.4.1. Objetivo general	3
1.4.2. Objetivos específicos	3
 Capítulo 2. Metodología	
2.1. Ecuación (ejemplo)	4
2.2. Figura (ejemplo)	4
 Capítulo 3. Resultados	
3.1. Tablas (ejemplo)	5
 Capítulo 4. Discusión	
 Capítulo 5. Conclusiones	
Literatura citada	8
Anexos	10

Lista de figuras

Figura

Página

- | | | |
|----|--------------------------|---|
| 1. | Logo de CICESE | 4 |
|----|--------------------------|---|

Lista de tablas

Tabla	Página
1. Esquema de codificación.	5

Capítulo 1. Introducción

El diseño computacional de proteínas ha emergido como una disciplina clave en la biología sintética y la ingeniería biomolecular, con un gran número de aplicaciones en diversas industrias. A través de modelos *in silico*, es posible predecir secuencias de aminoácidos que adopten estructuras estables y funcionales, lo que reduce significativamente el tiempo y costo asociados con los métodos experimentales tradicionales (Woolfson, 2021). Dicho enfoque ha sido utilizado para crear enzimas con nuevas actividades catalíticas, desarrollar proteínas terapéuticas con propiedades mejoradas y diseñar materiales biológicos con funciones específicas (Huang et al., 2016; Kuhlman et al., 2003; Langan et al., 2019). Sin embargo, el diseño de proteínas es un problema complejo que involucra la optimización de múltiples criterios, como la estabilidad estructural, la afinidad de unión y la especificidad funcional de las secuencias proteicas propuestas. Los algoritmos de optimización multiobjetivo han demostrado ser herramientas poderosas para abordar estos desafíos, permitiendo explorar grandes espacios de secuencias y encontrar soluciones que equilibran diferentes criterios de diseño (Deb, 2001). En este contexto, esta tesis se enfoca en el desarrollo y aplicación de algoritmos de optimización multiobjetivo para el diseño computacional de proteínas, con el objetivo de mejorar la eficiencia y efectividad del proceso de diseño y contribuir al avance de la biología sintética y la ingeniería biomolecular.

La introducción es la presentación del trabajo, informa sobre tres elementos muy importantes de la investigación: el propósito, la importancia del trabajo realizado y el conocimiento actual del tema. El texto debe comenzar con consideraciones generales y se recomienda terminar con el propósito del trabajo. Es conveniente que sea breve (no mayor de cinco páginas) y debe existir coherencia entre las distintas secciones que se presentan. La redacción debe ser clara, directa y sencilla, de tal forma que un lector no familiarizado con el tema pueda comprender el alcance del trabajo y motive su lectura.

1.1. Antecedentes

El diseño computacional de proteínas, enmarcado dentro del campo de la bioinformática, es una disciplina que usa métodos matemáticos y computacionales para predecir y diseñar secuencias de aminoácidos que adopten estructuras tridimensionales específicas con funciones biológicas deseadas (Huang et al., 2016). Muchas de las tareas que busca atender dicha disciplina – por ejemplo, el alineamiento de secuencias, la predicción de estructuras y el análisis y modelado de redes moleculares – pueden ser formuladas como problemas de optimización de gran complejidad, con rasgos como alta dimensionalidad, multimodalidad

y múltiples restricciones. (Zhang, 2009).

En esta sección se profundiza en el conocimiento acerca del tema y la relación con el trabajo de investigación. Es una descripción apoyada por la literatura citada. Sin tratar de resumir todo lo que se conoce del tema, ni de demostrar que se conoce toda la literatura. Hay que limitarse al tema específico del trabajo de investigación y a las contribuciones que se consideren más relevantes. (Cordero Esquivel, 1988)

1.2. Justificación (opcional)

En la justificación se debe expresar el por qué del estudio y la razón de su realización. Convencer al lector de que se hizo una investigación significativa: la importancia, la pertinencia del tema, el objeto de estudio y la utilidad de los resultados obtenidos.

1.3. Hipótesis (opcional)

La hipótesis del trabajo de tesis es la proposición que se pretende confirmar o refutar. No todas las investigaciones tienen hipótesis, sólo la necesitan aquellas que han rebasado la fase exploratoria. La hipótesis es la explicación que se le da a un hecho o fenómeno observado. Puede haber varias hipótesis para una misma pregunta de investigación y éstas no han de ser tomadas como verdaderas, sino que serán sometidas a pruebas para confirmar su veracidad.

1.4. Objetivos

Los objetivos son las metas del conocimiento que se pretenden alcanzar, a qué resultados se quiere llegar. Es decir, son el destino de la tesis; el marco teórico, el terreno y la metodología, el camino a seguir. Los objetivos deben expresarse en forma concisa, clara e inequívoca. Se expresan comenzando con un verbo en infinitivo, por ejemplo: analizar, comparar, definir, clasificar, por mencionar algunos. Los objetivos pueden dividirse en generales y específicos. El objetivo general es la descripción de la finalidad principal del estudio. Los objetivos específicos, si los hay, son considerados como secundarios. Son enunciados que

facilitan la comprensión de las metas. Los errores más comunes en la definición de los objetivos son:

- Ser demasiado amplios y generalizados.
- Objetivos específicos no contenidos en los generales.
- Planteo de pasos como si fueran objetivos (confundirlos con métodos o metas).
- Confusión entre objetivos y políticas o planes para llegar a lo que es la finalidad práctica.
- Falta de relación entre los objetivos, el marco teórico y la metodología.

1.4.1. Objetivo general

1.4.2. Objetivos específicos

Capítulo 2. Metodología

En esta sección se describen de forma detallada los procedimientos utilizados para la realización de la investigación, con el propósito de que se pueda reproducir. Incluye una descripción de los insumos utilizados por ejemplo: muestras colectadas, mediciones variables en el área de estudio o datos disponibles en alguna base de datos de acceso público o privado. Esta sección es apropiada para describir diseños experimentales, protocolos de adquisición e instrumentación empleados. Se escribe en tiempo pasado y no debe ser una lista de materiales ni de pasos a seguir y es conveniente evitar el uso de términos ambiguos tales como: frecuentemente, regularmente, aproximadamente.

2.1. Ecuación (ejemplo)

$$\theta(t) = \theta(0) + \int_0^t \theta(\tau) d\tau \quad (1)$$

2.2. Figura (ejemplo)



Figura 1. Logo de CICESE

El Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE) fue la segunda institución creada por el Consejo Nacional de Ciencia y Tecnología (CONACYT) para descentralizar las actividades científicas y tecnológicas en México". En la Figura 1 se puede ver el logo de CICESE.

Capítulo 3. Resultados

Esta es la sección más importante de la tesis. Se sugiere presentar los resultados en el mismo orden que la metodología. Se describen con la ayuda de tablas, gráficas o figuras, de manera que faciliten su comprensión. Las tablas se utilizan generalmente para comunicar valores concretos de los datos, mientras que las figuras son mejores para mostrar tendencias o relaciones entre variables. Se recomienda utilizar subtítulos para agrupar resultados similares y para separar resultados de diferentes parámetros. Una regla básica es no duplicar la información de una tabla en gráficas. Si los mismos resultados de una tabla o figura dentro del contenido de la tesis proporcionan información relevante, se recomienda ponerla como anexo.

3.1. Tablas (ejemplo)

Tabla 1. Esquema de codificación.

Categoría	Comportamiento	Definición
Actividad física. Los movimientos que el niño realiza durante el juego	Estacionario	Mantenerse inmóvil durante 3 segundos o más. Movimientos de dedos o pies
	Movimiento extremidades	Movimientos del tronco, brazos y piernas sin mover todo el cuerpo de un lugar a otro
	Translocacion lento	Mover el cuerpo de un lugar a otro con una velocidad lenta.
	Translocacion medio	Mover el cuerpo de un lugar a otro con una velocidad moderada.
	Translocacion rápido	Mover el cuerpo de un lugar a otro con una velocidad rápida

Capítulo 4. Discusión

En esta sección se deben interpretar y contrastar los resultados obtenidos. Se recomienda evitar demasiadas citas, ya que dificulta la lectura; hasta tres citas de fuentes formales son suficientes para respaldar cualquier afirmación. Es recomendable discutir las limitaciones del trabajo y de los métodos utilizados para minimizar o compensar esas limitaciones. Es importante enfatizar las implicaciones de los resultados y trazar futuras líneas de investigación.

Capítulo 5. Conclusiones

Esta sección refuerza lo expuesto en la introducción, se exponen de manera concisa los hallazgos más importantes de la investigación, contrastándolos con las hipótesis y objetivos planteados al inicio del trabajo. Se recomienda agregar las aportaciones, las limitaciones que se tuvieron y recomendaciones para trabajo futuro de la investigación. No se deben mencionar elementos no estudiados en la investigación ni repetir el contenido de la tesis.

Literatura citada

- APA (American Psychiatric Association) (2019). *Publication Manual of the American Psychological Association*, (11va ed.). Washington, DC.
- Brave, R. (2001). Governing the Genome. <http://online.sfsu.edu/rone/GEessays/GoverningGenome.html>.
- Burns, A. (2005). Action research: an evolving paradigm? *Language Teaching*. <https://doi.org/10.1017/S0261444805002661>.
- Cadena Ramírez, J. L. (2020). Informe anual de operaciones de embarcaciones menores del Departamento de Embarcaciones Oceanográficas durante 2012. (Informe técnico No. 3025), Centro de Investigación Científica y de Educación Superior de Ensenada, B.C.
- Cordero Esquivel, B. (1988). Evaluación de tres dietas artificiales para *Procambarus clarkii* (Girard): (Decapoda-cambaridae), en condiciones de laboratorio. [Tesis de Maestría en Ciencias, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California].
- Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK.
- Delgado Argote, L. A. & Sánchez Carballido, E. A. (1990). Análisis tectónico del sistema transpresivo neogénico entre Macuspana, Tabasco, y Puerto Ángel, Oaxaca. *Revista mexicana de ciencias geológicas*, 9(1), 21–32.
- Fainstein, S. S. & Campbell, S., editors (2011). *Readings in urban theory*, (3a ed.). Wiley-Blackwell, Cambridge.
- FAO Fisheries and Aquaculture Department (2006). Species Fact Sheets - *Paralichthys olivaceus* (Temminck & Schlegel, 1846).
- FAO (Food and Agriculture Organization) (2012). *The State of World Fisheries and Aquaculture 2012*. Rome.
- Favela, J., Tentori, M., Castro, L. A., Gonzalez, V. M., Moran, E. B., & Martínez-García, A. I. (2007). Activity Recognition for Context-aware Hospital Applications: Issues and Opportunities for the Deployment of Pervasive Networks. *Mobile Networks and Applications*, 12(2-3), 155–171. <https://doi.org/10.1007/s11036-007-0013-5>.
- Field, D. B., Baumgartner, T. R., Charles, C. D., Ferreira-Bartrina, V., & Ohman, M. D. (2006). Planktonic Foraminifera of the California Current Reflect 20th-Century Warming. *Science*, 311(5757), 63–66. <https://doi.org/10.1126/science.1116220>.
- García-Valdez, M., Trujillo, L., de Vega, F., Merelo Guervós, J. J., & Olague, G. (2013). EvoSpace: A Distributed Evolutionary Platform Based on the Tuple Space Model. In Esparcia-Alcázar, A. I., editor, *Applications of Evolutionary Computation*, Lecture Notes in Computer Science, 499–508, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Huang, P., Boyken, S. E., & Baker, D. (2016). The coming of age of de novo protein design. *Nature*, 537(7620), 320–327. <https://doi.org/10.1038/nature19946>.
- Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649), 1364–1368. <https://doi.org/10.1126/science.1089427>.

- Langan, R. A., Boyken, S. E., Ng, A. H., Samson, J. A., Dods, G., Westbrook, A. M., Nguyen, T. H., Lajoie, M. J., Chen, Z., Berger, S., Mulligan, V. K., Dueber, J. E., Novak, W. R. P., El-Samad, H., & Baker, D. (2019). De novo design of bioactive protein switches. *Nature*, 572(7768), 205–210. <https://doi.org/10.1038/s41586-019-1432-8>.
- Lloyd, C.-B. (2005). *Growing up global: The changing transitions to adulthood in developing countries*. National Academies Press.
- National Research Council and Institute of Medicine (2005). *Growing up global: The changing transitions to adulthood in developing countries*. National Academies Press.
- Ogata, K. (2010). *Modern Control Engineering*. Instrumentation and controls series. Prentice Hall, New York.
- Postgate, J. (1969). Viable counts and Viability. In Norris, J. & Ribbons, D., editors, *Methods in Microbiology*, (pp. 611–628). Academic Press, London y New York.
- South, J. & Blass, B. (2001). *The future of modern genomics*. Blackwell, London.
- Torres, X., Delgado-Argote, L., Avilez, P., & Weber, B. (2012). Geologic Analysis of the San Jerónimo pluton in the Southern part of the Peninsular Ranges Batholith, Baja California, México. In *Geological Society of America Cordilleran Section-108th Annual Meeting*, volume 1, 5–10. <https://doi.org/10.13140/2.1.4183.2649>.
- Treviño, L., Olague, G., Hammoud, R., & Hernandez, B. (2005). Automatic Feature Localization in Thermal Images for Facial Expression Recognition. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, volume 3, 15–14, San Diego, C.A. IEEE. <https://doi.org/10.1109/CVPR.2005.415>.
- Trujillo, L. & Olague, G. (2006). Synthesis of interest point detectors through genetic programming. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation - GECCO '06*, 887–894, New York, New York, USA. ACM Press. <https://doi.org/10.1145/1143997.1144151>.
- Woolfson, D. N. (2021). A brief history of de novo protein design: Minimal, rational, and computational. *Journal of Molecular Biology*, 433(20), 167160. From Protein Sequence to Structure at Warp Speed: How AlphaFold Impacts Biology, <https://doi.org/https://doi.org/10.1016/j.jmb.2021.167160>.
- Zhang, Y. (2009). Protein structure prediction: when is it useful? *Current Opinion in Structural Biology*, 19(2), 145–155. <https://doi.org/10.1016/j.sbi.2009.02.005>.

Anexos

Los anexos son secciones relativamente independientes que permiten conocer más a fondo aspectos específicos, que por su longitud o naturaleza no conviene incluir dentro del documento principal. Son elementos para dar una información más completa y que es útil para investigaciones futuras. Los anexos constituyen una sección adicional a la organización del trabajo y en ellos debe incluirse material complementario como: estadísticas, gráficas, fotografías, mapas, tablas, programas de cómputo, etcétera. Si no se menciona en el texto principal no deben incluirse. La información de los anexos debe ser completa, de manera que pueda utilizarse de forma independiente.

De acuerdo con las características de la información, el formato es libre. Se recomienda colocarlos en el orden en que están citados en el texto y, de preferencia, usando letras mayúsculas (Anexo A, Anexo B, Anexo C, etc.)

Las imágenes y las tablas que se añadan a los anexos deben de seguir con la numeración del contenido.