

Aspectos generales

Título:	Evolución y Diseño de Proteínas
Programas de posgrado o planes de estudio en donde se ofertará adicionalmente:	
PDCB, PMDCBQ, PCBIOL, LCG	
Área del conocimiento:	Bioquímica, biofísica y biología estructural
Semestre:	2027-1
Modalidad:	Tópico selecto
Horario:	Miercoles y Viernes (11:00 am - 1:00 pm)
No. sesiones:	32
Horas por sesión:	2.0
Total alumnos PDCB:	20
Total alumnos:	40
Videoconferencia:	Si
Lugar donde se imparte:	Centro de Ciencias Genómicas
Informes:	proteins-ai@ccg.unam.mx

Métodos de evaluación

MÉTODO	PORCENTAJE	NOTAS
Examen parcial	20%	
Exposición de artículo (journal club)	15%	
Participación y discusión	10%	
Proyecto integrador	35%	
Tareas y ejercicios computacionales	20%	

Contribución de este curso/tópico en la formación del alumnado del PDCB:

El alumno será capaz de analizar los mecanismos evolutivos que dan origen a la diversidad funcional de proteínas, y aplicar herramientas computacionales y experimentales de vanguardia para el diseño de proteínas con propiedades específicas, integrando conceptos de plegamiento, evolución molecular, inteligencia artificial y biología sintética.

Profesor (a) responsable

Nombre:	Farías Rico José Arcadio
Teléfono:	(777) 3291777 Ext. 38479
Email:	jose.arcadio.farias.rico@gmail.com

Profesores (as) participantes

PARTICIPANTE	ENTIDAD O ADSCRIPCIÓN	SESIONES
--------------	-----------------------	----------

FARIAS RICO JOSÉ ARCADIO
Responsable

Centro de Ciencias Genómicas

EXAMEN PARCIAL (Módulos 1–3)

- 1.1 El origen de la vida, las proteínas y el ribosoma
- 1.2 Estructura de proteínas: del aminoácido al dominio
- 1.4 Plegamiento cotraduccional y el papel del ribosoma
- 1.5 Evolución molecular de proteínas
- 2.1 Alineamientos y filogenias y clustering moleculares
- 2.2 Modelos de aprendizaje profundo para anotación de función en proteínas
- 2.3 Predicción de estructura con IA (AlphaFold, ESMFold)
- 2.4 Búsqueda estructural y minería (Foldseek, DALI)
- 3.1 Principios de evolución dirigida
- 3.2 Métodos de selección y screening (ribosome display)
- 3.4 Despliegue en levadura para selección de proteínas de unión
- 3.5 Casos de éxito y journal club
- 4.1 Diseño racional clásico
- 4.2 Diseño computacional con Rosetta (scoring function only)
- 4.3 Diseño inverso: ProteinMPNN y modelos de lenguaje
- 4.5 Diseño de interacciones y ensamblajes (BindCraft)
- 4.6 Taller práctico: pipeline de diseño computacional
- 5.1 Ingeniería de enzimas para biocatálisis
- 5.2 Proteínas terapéuticas (casos de éxito)
- 5.3 Biosensores y circuitos genéticos (AP-GFP)
- 5.4 Sistemas de expresión libres de células
- 5.5 Del diseño al laboratorio: validación experimental
- 6.1 IDPs, Ensembles
- 6.2 Claude para la ciencia de proteínas
- 6.3 Claude para la ciencia de proteínas 2
- 6.4 Presentaciones de proyectos integradores (I)
- 6.5 Presentaciones de proyectos integradores (II) y cierre

FERNÁNDEZ VELASCO DANIEL ALEJANDRO
Integrante

Facultad de Medicina

Plegamiento de proteínas

LAZCANO ARAUJO ANTONIO EUSEBIO
Integrante

Facultad de Medicina

1.1 El origen de la vida, las proteínas y el ribosoma

PASTOR COLON CARMEN NINA
Integrante

Instituto de Fisiología Celular

Dinámica Molecular

ROMERO ROMERO SERGIO
Integrante

Instituto de Fisiología Celular

Modelos generativos de diseño de proteínas

Introducción

1. Justificación

El diseño de proteínas ha experimentado una revolución sin precedentes con la llegada de herramientas de inteligencia artificial como AlphaFold, ProteinMPNN y RFdiffusion. Estos avances permiten hoy diseñar proteínas con un éxito experimental que era impensable hace una década. Sin embargo, la formación sistemática en estos temas es escasa en los programas de posgrado en México y América Latina. Este curso llena esa brecha formativa, ofreciendo a estudiantes de posgrado de la UNAM y de la Licenciatura en Ciencias Genómicas una visión integral que conecta los principios evolutivos fundamentales con las herramientas computacionales de vanguardia y las aplicaciones biotecnológicas emergentes. El curso está diseñado para ser válido en múltiples programas de posgrado de la UNAM (PDCB, PMDCBQ, PCBIOL) y como optativa para la LCG, cumpliendo con el estándar de 64 horas semestrales (4 h/semana, 16 semanas, 8 créditos).

2. Descripción general

Este curso ofrece una visión integral de cómo la evolución natural ha moldeado la diversidad de proteínas y cómo podemos aprovechar estos principios evolutivos, junto con herramientas computacionales de última generación, para diseñar proteínas con funciones novedosas. El curso abarca desde los orígenes de las proteínas y el ribosoma, pasando por los fundamentos de estructura, plegamiento (incluido el plegamiento cotraduccional) y evolución proteica, hasta las fronteras del diseño computacional con inteligencia artificial, incluyendo técnicas de evolución dirigida, diseño racional, modelos generativos, y aplicaciones en biología sintética.

El alumno integrará conocimientos teóricos con ejercicios computacionales prácticos y análisis crítico de la literatura primaria, culminando en un proyecto integrador donde propondrá una estrategia original de diseño proteico. El curso está diseñado tanto para estudiantes de posgrado (maestría y doctorado) como para alumnos avanzados de la Licenciatura en Ciencias Genómicas.

3. Objetivos del curso

Objetivo general:

El alumno será capaz de analizar los mecanismos evolutivos que dan origen a la diversidad funcional de proteínas, y aplicar herramientas computacionales y experimentales de vanguardia para el diseño de proteínas con propiedades específicas, integrando conceptos de plegamiento, evolución molecular, inteligencia artificial y biología sintética.

Objetivos específicos:

1. Comprender la relación entre secuencia, estructura, dinámica y función de proteínas en un contexto evolutivo, desde los orígenes de la vida hasta el plegamiento cotraduccional.
 2. Utilizar herramientas bioinformáticas y de modelado estructural (AlphaFold, ESM, Foldseek) para el análisis y diseño de proteínas.
 3. Evaluar críticamente las estrategias de evolución dirigida (incluidas aproximaciones guiadas por ML) y diseño racional/computacional.
 4. Aplicar pipelines de diseño computacional (RFdiffusion, ProteinMPNN, AlphaFold2) para generar y validar in silico proteínas de novo.
 5. Conocer las aplicaciones del diseño de proteínas en biocatálisis, terapia, biosensores y biología sintética.
 6. 7. Analizar las implicaciones éticas y de bioseguridad del diseño de proteínas con IA.
- Desarrollar una propuesta original de diseño proteico integrando métodos computacionales y experimentales.

Temario**6. Contenido temático detallado**

Módulo 1. Fundamentos: Orígenes, estructura, función y evolución de proteínas (10 horas – Semanas 1–3)

Objetivo: Establecer las bases conceptuales sobre los orígenes de las proteínas, su relación con el ribosoma y la evolución de la traducción, así como la estructura, plegamiento (incluido el cotraduccional) y evolución molecular de proteínas.

Contenido temático:

- 1.1 El origen de la vida, las proteínas y el ribosoma (2h) (Antonio Lazcano)
 - Péptidos prebióticos: síntesis abiótica de aminoácidos, activación y condensación
 - Cooperación péptidos-ácidos nucleicos: la etapa péptido-polinucleótido (Fried et al. 2022)
 - Roles funcionales de péptidos primitivos: catálisis, compartimentalización y estructura
 - Historia evolutiva del ribosoma: modelo de acreción (Petrov et al. 2015)
 - Fases de evolución ribosomal: del plegamiento de RNA a la proteinización
 - El ribosoma como fósil molecular: 4 mil millones de años de historia (Alva and Lupas)
- 1.2 Estructura de proteínas: del aminoácido al dominio (2h)
 - Repaso de niveles estructurales y fuerzas estabilizantes
 - Clasificaciones estructurales: SCOP2, CATH y ECOD
 - Relación secuencia-estructura-función: el dogma y sus excepciones
- 1.3 Plegamiento de proteínas: paisaje energético (2h) (Alejandro Fernandez)
 - Termodinámica del plegamiento: paradoja de Levinthal y paisaje energético (funnels)
 - Mecanismos de plegamiento: nucleación-condensación, framework y difusión-colisión
 - Chaperonas moleculares: GroEL/ES, Hsp70/DnaK, chaperonas del grupo II
- 1.4 Plegamiento cotraduccional y el papel del ribosoma (2h)
 - Vectorialidad del plegamiento in vivo: del N al C-terminal
 - El túnel de salida ribosomal como entorno de plegamiento
 - Force Profile Analysis (FPA) y estudios del grupo von Heijne
 - Implicaciones para el diseño de proteínas: codon usage y velocidad traduccional
- 1.5 Evolución molecular de proteínas (2h)
 - Divergencia, convergencia y evolución neutral
 - Duplicación génica, neofuncionalización y subfuncionalización
 - Familias de dominios: repeticiones (ankyrin, leucine-rich, WD40, ?-propellers)
 - Recombinación de dominios, shuffling exónico y fusión génica

Módulo 2. Herramientas computacionales para el análisis de proteínas (10 horas – Semanas 3–5)

Objetivo: Dominar las herramientas computacionales modernas para analizar secuencias, predecir estructuras, realizar búsquedas estructurales y simular dinámica de proteínas.

Contenido temático:

- 2.1 Alineamientos y filogenias moleculares (2h)
 - Métodos de alineamiento múltiple: MUSCLE, MAFFT, HHblits, MMseqs2
 - Modelos evolutivos y filogenias de familias proteicas
 - Perfiles HMM y bases de datos Pfam/InterPro
- 2.2 Modelos de aprendizaje profundo para anotación de función en proteínas (2h)
 - Modelos de lenguaje proteico (ESM, ProtTrans) y representaciones latentes de función
 - Predicción de función a partir de secuencia: GO terms, sitios activos, sitios de unión
 - Anotación funcional de proteínas huérfanas y el dark proteome
 - Integración con bases de datos: UniProt, InterPro, Pfam
- 2.3 Predicción de estructura con IA (2h)
 - AlphaFold2: arquitectura (Evoformer, Structure Module), MSA y templates
 - AlphaFold3: predicción de complejos, ácidos nucleicos y ligandos
 - ESMFold, RoseTTAFold, OpenFold: modelos de lenguaje proteico para estructura
 - Limitaciones y métricas de confianza (pLDDT, PAE)
- 2.4 Búsqueda estructural y minería de estructuras (2h)
 - DALI, TM-align y Foldseek: búsqueda por similitud estructural

- AlphaFold Protein Structure Database y ESM Atlas
- Structural BLAST y detección de relaciones evolutivas remotas

Página 7

UNAM – CCG – PDCB / PMDCBQ / PCBIOL / LCG

- Manejo de software molecular: PyMOL — visualización, alineamiento estructural y análisis de superficies

2.5 Simulación y dinámica molecular (2h) (Nina Pastor)

- Fundamentos de dinámica molecular (MD) all-atom y coarse-grained
- Enhanced sampling: replica exchange, metadynamics, steered MD
- Aplicaciones: estabilidad térmica, interacciones, alostería y plegamiento cotraduccional

Módulo 3. Evolución dirigida (10 horas – Semanas 6–8)

Objetivo: Comprender los fundamentos, las metodologías clásicas y las aproximaciones modernas (incluyendo ML-guided) de evolución dirigida como estrategia de ingeniería de proteínas.

Contenido temático:

3.1 Principios de evolución dirigida (2h)

- Mutagénesis aleatoria: error-prone PCR, mutator strains
- DNA shuffling y recombinación in vitro: StEP, RACHITT, SCHEMA
- Bibliotecas combinatorias y exploración del espacio de secuencias
- Mutagénesis por saturación y site-saturation mutagenesis (SSM)

3.2 Métodos de selección y screening (2h)

- Display en fago y ribosoma.
- Selección por FACS y microfluídica (droplet-based screening)
- Ensayos de alto rendimiento (HTS) y compartimentalización in vitro
- Selección negativa y contraselecciones genéticas

3.3 Evolución continua y automatizada (2h) (Donato Garcia)

- PACE (Phage-Assisted Continuous Evolution)
- OrthoRep: replicación ortogonal para evolución in vivo
- Evolución dirigida automatizada con robótica de laboratorio

3.4 Despliegue en levadura para selección de proteínas de unión (2h)

- Fundamentos del yeast surface display: sistema Aga2p y presentación en superficie
- Construcción y escrutinio de bibliotecas de variantes por citometría de flujo (FACS)
- Selección de binders: anticuerpos, nanobodies, DARPins y proteínas de repetición
- Integración con diseño computacional: enriquecimiento guiado por predicción de estructura

3.5 Casos de éxito y análisis crítico (2h)

- Premio Nobel 2018 (Frances Arnold): lecciones y principios
- Evolución de enzimas para química verde y biocombustibles
- Sesión de journal club: artículos seminales del campo

Página 8

UNAM – CCG – PDCB / PMDCBQ / PCBIOL / LCG

Módulo 4. Diseño racional y computacional de proteínas (12 horas – Semanas 9–11)

Objetivo: Integrar las estrategias de diseño racional, computacional y basado en IA para la creación de proteínas con funciones a la medida, incluyendo experiencia práctica con pipelines de diseño.

Contenido temático:

4.1 Diseño racional clásico y ASR (2h)

- Mutagénesis sitio-dirigida y análisis estructura-función
- Estabilización: puentes disulfuro, empaquetamiento del core, consensus design
- Ingeniería de especificidad y actividad enzimática
- Diseño de proteínas ancestrales (Ancestral Sequence Reconstruction, ASR)

4.2 Diseño computacional con Rosetta (2h)

- Funciones de energía y campos de fuerza (REF2015)
- Fixed-backbone design y flexible-backbone design
- Diseño de novo de plegamientos: TIM barrels, α -barrels, coiled-coils, β -proteins
- Rosetta Scripts y PyRosetta: flujos de trabajo prácticos

4.3 Diseño inverso: ProteinMPNN y modelos de lenguaje (2h)

- ProteinMPNN: arquitectura, entrenamiento y benchmark
- LigandMPNN y SolubleMPNN: variantes especializadas
- Comparación con métodos clásicos de diseño de secuencias

4.4 Modelos generativos de difusión (RFdiffusion, Chroma) (2h) (Sergio Romero)

- RFdiffusion: fundamentos y condicionamiento (motif scaffolding, binder design)
- Chroma, FrameDiff y Genie: comparación de arquitecturas
- Modelos de flujo (flow matching): FoldFlow, FrameFlow
- Pipeline completo: RFdiffusion $\rightarrow$ ProteinMPNN $\rightarrow$ AlphaFold2 (validación in silico)

4.5 Diseño de interacciones y ensamblajes (2h)

- Diseño de binders, mini-proteínas y péptidos
- Nano-anticuerpos sintéticos y proteínas de repetición (DARPins, Affibodies)
- Ensamblajes simétricos, jaulas proteicas y nanotecnología proteica
- Diseño de interfases proteína-DNA (diseño de integrasas y recombinasas)

4.6 Taller práctico: pipeline de diseño computacional (2h)

- Ejercicio guiado: diseñar un binder usando RFdiffusion + ProteinMPNN

- Validación in silico con AlphaFold2 y métricas de calidad
- Análisis de resultados y discusión de viabilidad experimental

Módulo 5. Aplicaciones en biología sintética y biotecnología (10 horas – Semanas 12–14)

Objetivo: Explorar las aplicaciones actuales y emergentes del diseño de proteínas en biocatálisis, medicina, biosensores y biología sintética, incluyendo consideraciones de validación experimental.

Contenido temático:

5.1 Ingeniería de enzimas para biocatálisis (2h)

- Diseño de enzimas con actividades no naturales (Kemp eliminasa, Diels-Alderasas)
- Cascadas enzimáticas y metabolismo sintético
- Enzimas termoestables y adaptadas a solventes no convencionales
- Aplicaciones industriales: química verde, biocombustibles, plásticos

5.2 Proteínas terapéuticas (2h)

- Ingeniería de anticuerpos: humanización, maduración de afinidad, fragmentos (scFv, Fab)
- Nanocuerpos (VHH) y anticuerpos biespecíficos
- Proteínas de diseño contra pandemias: vacunas y antivirales

5.3 Biosensores y circuitos genéticos basados en proteínas (2h)

- Biosensores fluorescentes: diseño de GFP circulares y FRET-based
- Switches alostéricos de diseño (allosteric protein switches)
- Circuitos lógicos proteicos y biología sintética celular
- Optogenética: proteínas fotoactivables de diseño

5.4 Sistemas de expresión libres de células (2h)

- Fundamentos de los sistemas de transcripción-traducción in vitro (TXTL)
- Plataformas: extractos de E. coli, trigo y reticulocitos; sistema PURE
- Aplicaciones: síntesis de proteínas tóxicas, incorporación de ncAA, prototipado rápido

Página 9

UNAM – CCG – PDCB / PMDCBQ / PCBIOL / LCG

- Integración con diseño computacional: validación acelerada de proteínas de novo

5.5 Del diseño al laboratorio: validación experimental (2h)

- Estrategias de validación experimental: expresión, purificación y caracterización
- Ensayos biofísicos: CD, DSF, ITC, SPR, SEC-MALS
- Cribado experimental de bibliotecas de diseño
- Tasa de éxito y estrategias para mejorarla

Módulo 6. Fronteras del campo y proyecto integrador (12 horas – Semanas 14–16)

Objetivo: Sintetizar los conocimientos del curso en un proyecto original de diseño proteico, explorar las fronteras del campo y discutir las implicaciones éticas del diseño de proteínas.

Contenido temático:

6.2 Agentes de AI para biología de proteínas Claude para la ciencia de proteínas (2h)

- Establecimiento de entornos de trabajo utilizando agentes de AI
- Práctica de docking molecular usando Claude (Agentic code agent de Anthropic)
- Automated protein design pipelines
- Limitaciones actuales y preguntas abiertas

6.3 Trabajando de manera segura con gentes de AI (2h)

- Continuación de aplicaciones de IA en ciencia de proteínas
- Anotación estructural de un genoma usando Claude
- Trabajando de manera segura con Agentes de inteligencia artificial.
- Sesión de asesoría grupal para proyectos integradores

6.4 Presentaciones de proyectos integradores (I) (2h)

- Propuesta original de diseño de proteína para resolver un problema específico
- Defensa oral ante panel de evaluación (20 min + 10 min preguntas por alumno)

6.5 Presentaciones de proyectos integradores (II) y cierre (2h)

- Continuación de presentaciones orales
- Retroalimentación entre pares y discusión final del curso
- Evaluación del curso y perspectivas futuras

Página 10

UNAM – CCG – PDCB / PMDCBQ / PCBIOL / LCG

7. Metodología de enseñanza

1. Clases teóricas expositivas con discusión interactiva (aproximadamente 50% del tiempo total).

2. Sesiones prácticas computacionales: ejercicios guiados con AlphaFold, ColabFold, Rosetta/PyRosetta, ProteinMPNN, RFdiffusion, Foldseek y herramientas de análisis de secuencias (aproximadamente 25%).

3. Club de revista (journal club): análisis crítico de artículos recientes de alto impacto en cada módulo. Cada alumno presentará al menos un artículo durante el semestre.

4. Seminarios invitados: investigadores activos en el campo compartirán perspectivas sobre problemas abiertos (se programarán 5 seminarios a lo largo del semestre).

5. Proyecto integrador: los alumnos desarrollarán una propuesta de diseño proteico, desde la identificación del problema hasta la estrategia experimental propuesta, incluyendo validación in silico con las herramientas aprendidas en el curso.

6. Tareas computacionales: ejercicios prácticos que refuercen los conceptos vistos en clase, con entrega de reportes que incluyan análisis crítico de resultados.

Bibliografía

9. Bibliografía

Bibliografía básica y complementaria:

1. Branden, C. & Tooze, J. (1999). Introduction to Protein Structure. 2nd Ed. Garland Science.
2. Kessel, A. & Ben-Tal, N. (2018). Introduction to Proteins: Structure, Function, and Motion. 2nd Ed. CRC Press.
3. 4. Petsko, G.A. & Ringe, D. (2004). Protein Structure and Function. New Science Press.
- Fried, S.D., Fujishima, K., Makarov, M., Cherepashuk, I. & Hlouchova, K. (2022). Peptides before and during the nucleotide world: an origins story emphasizing cooperation between proteins and nucleic acids. J. R. Soc. Interface, 19(187), 20210641.
- Página 11
- UNAM – CCG – PDCB / PMDCBQ / PCBIOL / LCG
5. Petrov, A.S., Gulen, B., Norris, A.M., Kovacs, N.A., Bernier, C.R., Lanier, K.A., Fox, G.E., Harvey, S.C., Wartell, R.M., Hud, N.V. & Williams, L.D. (2015). History of the ribosome and the origin of translation. Proc. Natl. Acad. Sci. USA, 112(50), 15396–15401.
6. Arnold, F.H. (2018). Directed Evolution: Bringing New Chemistry to Life. Angew. Chem. Int. Ed., 57(16), 4143–4148.
7. Huang, P.S., Boyken, S.E. & Baker, D. (2016). The coming of age of de novo protein design. Nature, 537, 320–327.
8. Dauparas, J. et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. Science, 378(6615), 49–56.
9. Watson, J.L. et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature, 620, 1089–1100.
10. Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589.
11. Abramson, J. et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630, 493–500.
12. Notin, P. et al. (2024). Machine learning for functional protein design. Nature Biotechnology, 42, 216–228.
13. Pan, X. & Kortemme, T. (2021). Recent advances in de novo protein design. Curr. Opin. Struct. Biol., 68, 63–70.
14. Lutz, S. & Iamurri, S.M. (2018). Protein Engineering: Past, Present, and Future. Methods Mol. Biol., 1685, 1–12.
15. Waudby, C.A., Dobson, C.M. & Christodoulou, J. (2019). Nature and regulation of protein folding on the ribosome. Trends Biochem. Sci., 44(11), 914–926.
16. Coventry, B. & Baker, D. (2021). Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. PLOS Comp. Biol., 17(3), e1008061.

Nota: Se asignarán artículos adicionales de investigación primaria a lo largo del curso, seleccionados de revistas como Nature, Science, PNAS, Nature Biotechnology, Cell Systems, Current Opinion in Structural Biology y Annual Review of Biochemistry.

Observaciones

7. Metodología de enseñanza

1. Clases teóricas expositivas con discusión interactiva (aproximadamente 50% del tiempo total).
2. Sesiones prácticas computacionales: ejercicios guiados con AlphaFold, ColabFold, Rosetta/PyRosetta, ProteinMPNN, RFdiffusion, Foldseek y herramientas de análisis de secuencias (aproximadamente 25%).
3. Club de revista (journal club): análisis crítico de artículos recientes de alto impacto en cada módulo. Cada alumno presentará al menos un artículo durante el semestre.
4. Seminarios invitados: investigadores activos en el campo compartirán perspectivas sobre problemas abiertos (se programarán 5 seminarios a lo largo del semestre).
5. Proyecto integrador: los alumnos desarrollarán una propuesta de diseño proteico, desde la identificación del problema hasta la estrategia experimental propuesta, incluyendo validación in silico con las herramientas aprendidas en el curso.
6. Tareas computacionales: ejercicios prácticos que refuercen los conceptos vistos en clase, con entrega de reportes que incluyan análisis crítico de resultados.

Nota para alumnos de licenciatura (LCG):

Los alumnos de licenciatura recibirán apoyo adicional en los fundamentos de bioinformática estructural. La evaluación del proyecto integrador considerará el nivel académico del alumno. Se recomienda haber cursado Bioquímica, Biología Molecular y al menos una optativa de bioinformática.