

PROPUESTA DE PROYECTO DE INVESTIGACIÓN

IDENTIFICACIÓN DEL PROYECTO

UNIDAD:	Cuajimalpa.
DIRECCIÓN O DIVISIÓN:	Ciencias Naturales e Ingeniería.
DEPARTAMENTO:	Ciencias Naturales
ÁREA O SECCIÓN:	
NOMBRE DEL PROYECTO:	Diseño Computacional de Moduladores Biomoleculares y Materiales Funcionales Mediante Métodos Cuánticos y Aprendizaje Automático.
RESPONSABLE (Nombre y núm. eco.):	Dr. Felipe Aparicio Platas, 27126

DATOS DEL PROYECTO

FECHA DE APROBACIÓN:	16 de octubre de 2028
NÚMERO DE SESIÓN DE CONSEJO DIVISIONAL:	
FECHA DE INICIO:	17 de octubre de 2025
FECHA DE TÉRMINO:	16 de octubre de 2028
LÍNEA DE INVESTIGACIÓN:	Físicoquímica y Diseño Molecular
PROBLEMAS NACIONALES:	1. Salud. 2. Educación 3. Consumo sustentable de energía y desarrollo y aprovechamiento de energías renovables limpias.
VINCULACIÓN CON SECTORES SOCIALES:	Salud, educación y energía.
ÁREA DE CONOCIMIENTO:	1. Física y Ciencias de la Tierra. 2. Biología y Química.
ORIENTACIÓN:	Investigación Básica (X) Investigación Aplicada (X)

PARTICIPANTES EN EL PROYECTO

NOMBRE:	Dr. Roberto Bernal Jaquez.
NÚMERO ECONÓMICO:	33676
TABULADOR*:	
TIEMPO DE DEDICACIÓN:	4 horas/semana.
LÍNEA DE INVESTIGACIÓN:	Redes Neuronales y Machine Learning.
TIPO DE PARTICIPACIÓN:	Colaborador.

NOMBRE:	Dr. Gerardo Pérez Hernández.
NÚMERO ECONÓMICO:	31266
TABULADOR*:	
TIEMPO DE DEDICACIÓN:	4 horas/semana.
LÍNEA DE INVESTIGACIÓN:	Estudios de biomoléculas con métodos de dinámica molecular clásica.
TIPO DE PARTICIPACIÓN:	Colaborador.

NOMBRE:	Dr. Salomón de Jesús Alas Guardado.
NÚMERO ECONÓMICO:	28845
TABULADOR*:	
TIEMPO DE DEDICACIÓN:	4 horas/semana.
LÍNEA DE INVESTIGACIÓN:	Estudios de la estabilidad térmica de biomoléculas con métodos de dinámica molecular clásica.
TIPO DE PARTICIPACIÓN:	Colaborador.

* Si aplica.

NOMBRE:	Dr. Ricardo Romero Ochoa
NÚMERO ECONÓMICO:	
TABULADOR*:	
TIEMPO DE DEDICACIÓN:	4 horas/semana.
LÍNEA DE INVESTIGACIÓN:	Aplicación de machine learning para la búsqueda de moléculas con interés biológico.
TIPO DE PARTICIPACIÓN:	Colaborador.

NOMBRE:	Dr. Jorge Alberto Ramírez Cano (Universidad Veracruzana).
NÚMERO ECONÓMICO:	
TABULADOR*:	
TIEMPO DE DEDICACIÓN:	4 horas/semana.
LÍNEA DE INVESTIGACIÓN:	Aplicación de machine learning para la búsqueda de moléculas con interés tecnológico.
TIPO DE PARTICIPACIÓN:	Colaborador.

ALINEACIÓN AL PLAN DE DESARROLLO INSTITUCIONAL 2011–2024

CLAVE DE PROYECTO:

PROGRAMA INSTITUCIONAL: Investigación y formación de recursos humanos de alto nivel; fortalecimiento de la planta académica y de la infraestructura para la investigación.

PROYECTO OPERATIVO: Fortalecer líneas de investigación estratégicas mediante proyectos interdisciplinarios orientados a la solución de problemas nacionales en ciencia y tecnología, incluyendo la renovación y consolidación de cuerpos académicos y el uso eficiente de infraestructura científica.

ACCIÓN:

- 1) Desarrollar y ejecutar un protocolo de simulación molecular que integre a la Química Cuántica, a la Biología Cuántica y a la Inteligencia Artificial (Machine Learning).
- 2) Formar estudiantes (L–M–D) mediante tesis y estancias de investigación.
- 3) Publicar en revistas indexadas y promover acceso abierto y repositorio institucional; difundir resultados y vincular con sector público/industrial.

OBJETIVO ESTRATÉGICO: Realizar investigación que contribuya al desarrollo científico y tecnológico del país, con impacto social y formación de profesionales de alta calidad, articulando docencia, investigación y vinculación.

FACTOR CLAVE: Capacidades académicas (cuerpos académicos consolidados), infraestructura y equipamiento científicos, colaboración interinstitucional y vinculación con sectores productivos; gestión eficaz con metas, indicadores y evaluación continua.

INDICADOR:

Productos y resultados verificables por periodo:

- Artículos JCR/Scopus aceptados (meta anual).
- Proyectos terminales de licenciatura y tesis de posgrado concluidas (L–M–D) y estudiantes en formación.
- Presentaciones en congresos nacionales e internacionales y cursos/talleres impartidos.
- Protocolos/software/datasets depositados en repositorio institucional.
- Proyectos con financiamiento externo y convenios de vinculación.

Diseño Computacional de Moduladores Biomoleculares y Materiales Funcionales Mediante Métodos Cuánticos y Aprendizaje Automático

Dr. Felipe Aparicio Platas

Resumen

Este proyecto pretende desarrollar una plataforma integrada de diseño molecular que combine cálculos cuánticos, de un alto nivel de precisión, con algoritmos de aprendizaje automático o *machine learning* (ML), para diseñar y optimizar compuestos con propiedades específicas, tales como fármacos, inhibidores de canales iónicos o materiales con propiedades ópticas de interés tecnológico. La estrategia central consiste en calcular descriptores moleculares cuánticos mediante la Teoría de Funcionales de la Densidad (DFT) y su versión dependiente del tiempo (TD-DFT), utilizar estos descriptores para entrenar modelos predictivos físico-informados y aplicar optimización activa para explorar eficientemente el espacio químico. Los sistemas objetivo incluyen moduladores de canales iónicos con potencial terapéutico en neurociencia y cardiología, así como cromóforos orgánicos para diseñar materiales con propiedades ópticas novedosas. El proyecto generará candidatos moleculares validados computacionalmente, metodologías reproducibles codificadas como software de acceso abierto (open-source) y bases de datos curadas (datasets) disponibles públicamente.

1. PROPUESTA DE PROYECTO

1.1. Antecedentes

El descubrimiento tradicional de fármacos y materiales mediante cribado experimental enfrenta limitaciones fundamentales de costo, tiempo y cobertura del espacio químico. El diseño computacional racional ofrece una alternativa transformadora al permitir la predicción de propiedades moleculares y la optimización dirigida por objetivos múltiples [1]. Los métodos cuánticos proporcionan descripciones precisas de la estructura electrónica y la reactividad de moléculas y materiales, sin requerir datos experimentales previos, pero su costo computacional limita su aplicación directa a exploración extensiva [2]. Por otra parte, el aprendizaje automático o *machine learning* (ML) puede predecir propiedades a velocidades de milisegundos por molécula, aunque típicamente requiere grandes datasets de entrenamiento y carece de sesgo físico y químico explícito [3].

La integración sinérgica de estos enfoques permite superar estas limitaciones: generamos datasets de alta calidad mediante cálculos cuánticos selectivos, entrenamos modelos físico-informados que capturan relaciones estructura-propiedad fundamentales, y aplicamos estos modelos para explorar regiones vastas del espacio químico. El aprendizaje activo cierra el ciclo mediante selección inteligente de nuevas moléculas para cálculo, maximizando el aprendizaje con recursos computacionales mínimos [4].

Por otra parte, en el campo de diseño de fármacos, los curcuminoides (curcumina y sus derivados) representan candidatos prometedores como moduladores de canales iónicos. La curcumina natural

bloquea canales Kv1.4 con IC_{50} de aproximadamente $15 \mu M$ y modula la cinética de inactivación de Kv2.1 mediante mecanismos alostéricos [5, 6]. Sin embargo, presenta limitaciones de biodisponibilidad, fotodegradación y selectividad de subtipo que pueden superarse mediante diseño racional de análogos [7]. Los canales iónicos son targets terapéuticos relevantes para epilepsia, arritmias cardíacas y dolor crónico, entre otros, con mercados globales multimillonarios para el desarrollo de nuevas terapias, pero se requieren moduladores más selectivos [8].

En ciencia de materiales, los cromóforos orgánicos para captación de luz deben satisfacer requisitos simultáneos de absorción espectral apropiada, separación eficiente de carga, estabilidad fotoquímica y procesabilidad. El diseño tradicional mediante síntesis iterativa es lento y costoso. Los sistemas biológicos de fotosíntesis logran eficiencias cercanas al 100 % mediante arquitecturas moleculares optimizadas evolutivamente [9, 10]. Aunque la relevancia funcional de fenómenos cuánticos como la coherencia cuántica aún es motivo de debate, los principios estructurales subyacentes (disposición geométrica, acoplamiento excitónico) son transferibles al diseño de materiales sintéticos [11].

Las arquitecturas contemporáneas de ML molecular incluyen representaciones basadas en strings (SMILES y SELFIES), embeddings de subestructuras (Mol2Vec), y redes neuronales gráficas que operan sobre grafos moleculares [12–14]. La incorporación explícita de información física y química mediante descriptores cuánticos, simetrías incorporadas en arquitecturas, y funciones de pérdida con restricciones fisicoquímicas mejora sustancialmente la precisión predictiva, especialmente en regímenes de pocos datos y al extrapolar fuera del dominio de entrenamiento [2, 15].

La articulación de estos tres frentes –cálculos cuánticos, reactividad biomolecular e inteligencia artificial– es una oportunidad para construir modelos predictivos que cumplan con principios fisicoquímicos y aceleren el diseño racional de moduladores con potencial biomédico y en ciencia de materiales.

1.2. Objetivo general

- Desarrollar y validar una plataforma de diseño computacional que integre cálculos cuánticos de precisión con aprendizaje automático fisico-informado para diseñar y descubrir moduladores de canales iónicos con selectividad mejorada y cromóforos orgánicos para materiales fotovoltaicos eficientes.

1.3. Objetivos específicos

El proyecto persigue seis objetivos específicos mutuamente acoplados:

1. Caracterizar cuánticamente, mediante DFT/TD-DFT, un conjunto de al menos 500 compuestos (curcuminoides, cromóforos, etc.) para generar descriptores moleculares de estructura electrónica, reactividad química y propiedades espectroscópicas.
2. Modelar coherencias y entrelazamientos cuánticos efectivos en fotosistemas (FMO, LHC-II) y criptocromos (dCRY, ICRY) mediante ecuaciones maestras (Lindblad/Redfield) parametrizadas a partir de QM/MM, y contrastarlas con métricas observables.
3. Cuantificar energías libres de unión y modos de interacción para curcuminoides con canales iónicos objetivo mediante simulaciones de dinámica molecular y cálculos QM/MM, estableciendo relaciones estructura-actividad cuantitativas.

4. Desarrollar arquitecturas ML híbridas que combinen representaciones estructurales con descriptores cuánticos, incluyendo modelos predictivos de regresión y modelos generativos para diseño *de novo*.
5. Implementar pipelines de aprendizaje activo que optimicen simultáneamente afinidad, selectividad, estabilidad y propiedades ADMET mediante algoritmos de frontera de Pareto.
6. Validar predicciones mediante comparación con datos experimentales, desarrollar workflows reproducibles y datasets conocidos como localizables, accesibles, interoperables y reutilizables (FAIR), y publicar resultados en revistas de alto impacto.

1.4. Hipótesis

Proponemos que modelos ML incorporando descriptores cuánticos explícitos mostrarán precisión predictiva superior en más del 20 % comparados con métodos puramente estructurales, especialmente en regímenes de menos de 1000 ejemplos de entrenamiento. Las estrategias de aprendizaje activo que combinan incertidumbre epistémica con diversidad química reducirán en más del 50 % el número de cálculos cuánticos requeridos para identificar candidatos óptimos comparado con muestreo aleatorio. El diseño guiado por descriptores cuánticos, particularmente índices de reactividad y mapas electrostáticos, permitirá optimizar selectividad entre subtipos de canales iónicos en un factor superior a 10 manteniendo afinidad. Finalmente, los modelos entrenados en curcuminoides mostrarán transferibilidad significativa a scaffolds químicos distintos, validando la captura de principios fisicoquímicos generalizables.

2. METODOLOGÍA

2.1. Resumen de métodos a utilizar

El flujo de trabajo comprende cuatro métodos interconectados (módulos) que operarán de forma iterativa:

- El módulo de cálculos cuánticos genera datasets de propiedades moleculares mediante DFT/TD-DFT.
- El módulo de simulaciones moleculares cuantifica interacciones proteína-ligando mediante dinámica molecular clásica (MD) y modelos híbridos (QM/MM).
- El módulo de aprendizaje automático construye modelos predictivos y generativos físico-informados.
- El módulo de optimización activa selecciona adaptativamente candidatos para validación computacional costosa, refinando continuamente los modelos.

2.2. Caracterización cuántica

Calcularemos propiedades electrónicas y espectroscópicas para al menos 200 curcuminoides (naturales de COCONUT/NPATlas más derivados sintéticos con modificaciones sistemáticas en carbonilos, anillos aromáticos y espaciador central) y al menos 300 cromóforos (porfirinas, ftalocianinas, BO-DIPY, cianinas) mediante el funcional B3LYP-D3(BJ) con conjunto de base def2-TZVP y modelo de

solvente continuo SMD. Las geometrías se optimizarán hasta convergencia estricta (fuerza máxima < 0.0001 Hartree/Bohr).

Para cada molécula calcularemos orbitales frontera HOMO/LUMO con energías y densidades, índices de reactividad DFT conceptual (electronegatividad, dureza, electrofilia), funciones de Fukui para predicción de sitios reactivos, cargas parciales Hirshfeld, y momentos multipolares. Los espectros electrónicos se calcularán mediante TD-DFT con el funcional CAM-B3LYP optimizado para estados de transferencia de carga, obteniendo 10-20 estados excitados con longitudes de onda, fuerzas de oscilador y momentos de transición. Calcularemos también energías de reorganización nuclear mediante optimización de geometrías de estados excitados.

Estos cálculos generarán más de 50 descriptores cuánticos por molécula incluyendo gap HOMO-LUMO, potenciales de ionización/afinidad electrónica, índices de aromaticidad, superficies de potencial electrostático, volúmenes moleculares y descriptores topológicos. Validaremos la metodología comparando espectros predichos contra datos experimentales disponibles, con objetivo de $R^2 > 0.90$ y error absoluto medio < 20 nm en longitudes de onda.

2.3. Energética molecular

Prepararemos sistemas proteína-ligando, utilizando estructuras cristalográficas cuando estén disponibles (Kv1.2 de PDB) o modelos de homología/AlphaFold2 validados experimentalmente para canales sin estructura (Nav, Cav). Los sistemas se embeberán en bicapas lipídicas POPC, solvatarán con TIP3P y neutralizarán con NaCl 150 mM, resultando en sistemas de aproximadamente 150,000-300,000 átomos.

Realizaremos simulaciones MD con AMBER ff19SB para proteínas y GAFF2 para ligandos, siguiendo protocolos de equilibración graduales antes de producción de 100-500 ns por réplica con cinco réplicas independientes. Las simulaciones correrán a 310 K y 1 bar usando integradores leap-frog con paso de 2 fs. Calcularemos energías de unión mediante MM/PBSA sobre 1000 snapshots equiespaciados, obteniendo descomposiciones de contribuciones electrostáticas, Van Der Waals y solvatación.

Para compuestos prometedores calcularemos perfiles de energía libre mediante Umbrella sampling a lo largo de coordenadas de reacción relevantes (20-30 ventanas, 20-50 ns cada una, análisis WHAM) y metadinámica well-tempered para explorar múltiples modos de unión. Las transformaciones alquímicas mediante FEP/TI en 21-41 ventanas λ con soft-core potentials cuantificarán diferencias de energía libre entre análogos. Cálculos QM/MM con B3LYP-D3/def2-SVP para regiones QM de ≈ 150 átomos (ligando más residuos críticos) caracterizarán interacciones donde polarización electrónica es relevante.

2.4. Aprendizaje automático físico-informado

Construiremos representaciones moleculares híbridas combinando embeddings estructurales (Mol2Vec de 300 dimensiones, Morgan fingerprints de 2048 bits, grafos para GNNs) con vectores de 50 dimensiones de descriptores cuánticos normalizados. Implementaremos modelos de regresión incluyendo Random Forest como baseline interpretable, Gradient Boosting (XGBoost/LightGBM) con tuning extensivo, Gaussian Processes con kernels físico-informados para cuantificación de incertidumbre, redes neuronales densas de 3-5 capas con dropout y batch normalization, y GNNs tipo MPNN con agregación attention-weighted. Desarrollaremos representaciones moleculares físico-informadas usando métricas de distancia de Mahalanobis, en espacios latentes con covarianza informada por similitud de propiedades cuánticas.

Las funciones de pérdida incorporarán términos de datos, restricciones físicas (violaciones de ordenamiento HOMO<LUMO, simetrías) y regularización, ponderados por hiperparámetros optimizados. Entrenaremos mediante cross-validation 5-fold estratificado asegurando distribución balanceada de scaffolds, con ensembles de 10-20 modelos para robustez. Aplicaremos *transfer learning* pre-entrenando en datasets grandes (QM9, ZINC) antes de *fine-tuning* en datos específicos, y *multi-task learning* para predicción simultánea de propiedades correlacionadas.

Para diseño *de novo* implementaremos VAEs condicionales que mapean SMILES a espacio latente continuo de 256 dimensiones, condicionados en targets de propiedades objetivo. El decoder reconstruye SMILES válidos, permitiendo interpolación y muestreo dirigido. Implementaremos también *reinforcement learning* con políticas π que navegan el espacio químico maximizando recompensas multiobjetivo ponderadas (afinidad, selectividad, síntesis factible, drug-likeness) usando *Proximal Policy Optimization*.

La interpretabilidad se evaluará mediante SHAP values identificando características críticas, LIME para explicaciones locales, visualización de *attention weights* en GNNs, y análisis de *counterfactual explanations* mostrando modificaciones mínimas para alcanzar propiedades objetivo.

2.5. Optimización activa

Implementaremos optimización bayesiana con procesos gaussianos como modelos subrogados, utilizando kernels físico-informados que combinan similitud de Mahalanobis con similitud de descriptores cuánticos. Las *acquisition functions* incluirán *Upper Confidence Bound* balanceando exploración/explotación, *Expected Improvement* maximizando mejora esperada sobre mejor observación, y *Max-value Entropy Search* para multiobjetivo optimizando frontera de Pareto.

La cuantificación de incertidumbre distinguirá componentes epistémicas (ensambles, dropout bayesiano, varianza de GP) y aleatorias (ruido inherente), priorizando regiones con alta incertidumbre epistémica. El diversity sampling mediante MaxMin distance y k-center clustering en espacios de fingerprints y descriptores cuánticos evitará redundancia. Estableceremos jerarquía multi-fidelidad con predicción ML rápida para screening masivo de millones de compuestos, DFT/def2-SVP para refinamiento de top-1000, DFT/def2-TZVP para validación de top-100, y MD/QM-MM para caracterización final de top-20, aprendiendo correlaciones entre niveles para corregir predicciones de baja fidelidad.

El pipeline operará en ciclos: inicialización mediante muestreo diverso de 100 moléculas con cálculos cuánticos y entrenamiento baseline, seguido por iteraciones de batches de 20-50 moléculas donde acquisition functions identifican candidatos de librerías virtuales, filtros eliminan compuestos violando drug-likeness o conteniendo subestructuras problemáticas, cálculos selectivos validan candidatos, re-entrenamiento actualiza modelos, y análisis de convergencia determina terminación cuando mejora de hypervolume de Pareto es menor al 5 %.

2.6. Casos de estudio

Diseñaremos moduladores de Kv1.4 con $IC_{50} < 5 \mu M$, selectividad $> 20\times$ sobre Kv1.5, half-life fotoquímica > 4 h, y propiedades drug-like (cLogP 2-4, PSA $< 120 \text{ \AA}^2$). Caracterizaremos cuánticamente 150 curcuminoides, simularemos 30 con Kv1.4/Kv1.5, entrenaremos modelos predictivos, ejecutaremos 5 ciclos de optimización activa de 30 moléculas, y validaremos top-10 con MD extendido de 500 ns y QM/MM. El éxito requiere ≥ 3 candidatos satisfaciendo criterios, predicciones con MAE < 0.5 kcal/mol, y reducción 50 % de cálculos versus screening exhaustivo.

Para cromóforos de gap estrecho optimizaremos absorción 650-750 nm, gap 1.5-1.8 eV, HOMO - 5.2 a -5.6 eV, fuerza de oscilador >1.0 , y reorganización <0.3 eV. Calcularemos TD-DFT de 200 cromóforos base, entrenaremos VAE condicional, generaremos 1000 candidatos con restricciones, validaremos top-100 con DFT refinado, y analizaremos estabilidad fotoquímica. El éxito requiere ≥ 10 estructuras novedales con $> 50\%$ sintéticamente accesibles y predicción espectral con error < 30 nm.

Evaluaremos transferibilidad entrenando en curcuminoides y aplicando a chalconas estructuralmente relacionadas, comparando precisión versus modelo entrenado directamente en chalconas, identificando descriptores cuánticos más transferibles.

2.7. Formación de recursos humanos

El proyecto contempla la participación y formación de estudiantes en diferentes niveles académicos:

- **Cuatro** estudiantes de licenciatura para realizar servicio social. El proyecto permite incorporar estudiantes de los 4 programas de licenciatura que se ofrecen en la División de Ciencias Naturales e Ingeniería.
- **Tres** estudiantes de licenciatura para realizar Proyectos Terminales (PT-I, PT-II y PT-III). El proyecto permite incorporar estudiantes de los 4 programas de licenciatura que se ofrecen en la División de Ciencias Naturales e Ingeniería.
- **Dos** estudiantes candidatos a la maestría en el programa de Posgrado en Ciencias Naturales e Ingeniería (PCNI).
- **Un** estudiante candidato al doctorado en el programa de Posgrado en Ciencias Naturales e Ingeniería (PCNI).
- **Un** investigador posdoctoral financiado por la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti).

Como estrategia de continuidad académica, se ofrecerá a los estudiantes de licenciatura participantes la oportunidad de continuar su trayectoria en los programas de posgrado de nuestra institución, fomentando así la consolidación de líneas de investigación y el desarrollo de carrera científica dentro de la UAM.

También, en este proyecto se contempla la organización de un taller intensivo (workshop) sobre ML aplicada a la química (40 h), simulaciones moleculares (60 h), métodos cuánticos (30 h), y Scientific Python (20 h).

2.8. Productos de trabajo esperados

Se espera obtener los siguientes productos de trabajo:

- **Tres** artículos en revistas indizadas: un marco unificado para dinámica de coherencia en complejos de captación de luz, un estudio de caso sobre curcuminoides y canales iónicos usando IA *físico-informada* y un trabajo metodológico sobre aprendizaje activo multiobjetivo en diseño molecular.

- Se liberará **un** paquete de software: una librería de embeddings cuánticos con un módulo de optimización activa con soporte multiobjetivo.
- Se publicará **una** datasets FAIR, con propiedades cuánticas de cromóforos, acompañada de documentación y cuadernos reproducibles.
- **Dos** artículos de divulgación de la ciencia: uno para difundir el vínculo de la IA y la mecánica cuántica y otro para difundir la aplicaciones generadas en el proyecto.
- Participación en al menos 3 congresos nacionales y 2 congresos internacionales.
- Presentación de al menos 1 conferencias de divulgación, por año, para exponer los resultados obtenidos.
- Desarrollo de al menos 2 talleres para nivel medio superior o superior, donde se enseñen los fundamentos de la metodología utilizada en este proyecto.

2.9. Impacto esperado

En el plano metodológico, se espera establecer un *blueprint* replicable para IA *físico-informada* en química y biología computacional, con mejoras medibles en precisión y en capacidad de generalización. En lo aplicado, la plataforma permitirá priorizar 10–15 candidatos moduladores con perfiles superiores de afinidad y selectividad, con implicaciones potenciales en neurociencia y cardiología, y generará lineamientos transferibles a materiales orgánicos fotoactivos. Institucionalmente, el proyecto fortalecerá la línea de Fisicoquímica y Diseño Molecular, consolidará redes de colaboración y posicionará a la UAM-C en la intersección entre biología cuántica e inteligencia artificial.

2.10. Infraestructura computacional y reproducibilidad

El proyecto requerirá horas significativas de CPU para QM/MM y MD, y de GPU para el entrenamiento de modelos. Se utilizarán contenedores (Docker/Apptainer) con dependencias fijadas, *workflows* declarativos (Snakemake/CWL), control de versiones con Git y *testing* unitario e integral. Los datos y modelos se publicarán con DOIs y metadatos completos, asegurando que un equipo externo pueda reproducir los resultados de punta a punta.

Toda esta infraestructura computacional está disponible en el Laboratorio de Fisicoquímica y Diseño Molecular de la UAM-C.

2.11. Recursos requeridos

No se requiere financiamiento para este proyecto, toda vez que el responsable del mismo tiene acceso a horas de cómputo en el clúster YOLTLA instalado en el Laboratorio de Supercómputo y Visualización en Paralelo de la UAM-I. Además, contamos con 3 servidores de cómputo de alto rendimiento en el Laboratorio de Fisicoquímica y Diseño Molecular de la UAM-C, con disponibilidad de tiempo de CPU y GPU ilimitado.

La totalidad del software que usaremos en este proyecto será de código abierto (GROMACS/NAMD, ORCA/NWChem/CP2K, PyTorch/JAX, RDKit/OpenMM), con contenedores disponibles para descarga.

3.2. Información para el seguimiento del proyecto

PRODUCTO	AÑO 1	AÑO 2	AÑO 3
Formación de recursos humanos a nivel licenciatura			
Servicio Social	1	1	1
Proyectos Terminales	1	1	1
Formación de recursos humanos a nivel posgrado			
Maestría		2	
Doctorado			1
Pos-doctorado		1	
Publicaciones			
Artículos	1	1	1
Capítulos de libro			1
Memorias de congreso		1	1
Software libre y bases de datos curadas			2
Divulgación y Difusión			
Congresos	1	2	2
Conferencias	1	1	2
Talleres		1	1

4. Divulgación de la ciencia e impacto social

Además de los artículos de divulgación propuestos, el proyecto incorporará una estrategia de comunicación científica con públicos diferenciados. Para estudiantes de licenciatura y posgrado se ofrecerá un curso de 40 horas titulado “*Biología cuántica y aprendizaje de máquina*”, con materiales en español (videos, diapositivas y cuadernos ejecutables) liberados bajo licencia abierta. Para docentes de bachillerato se organizarán dos talleres centrados en conceptos de coherencia cuántica y diseño molecular ilustrados con simulaciones sencillas y analogías visuales. Para público general se realizará una serie de seminarios en línea y cápsulas de 10 minutos sobre preguntas guía (*¿qué significa coherencia cuántica en un sistema vivo? ¿cómo “aprende” una máquina química?*), difundidas en canales institucionales y redes sociales.

Referencias

- [1] Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018). doi: 10.1021/acscentsci.7b00572
- [2] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet: A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018). <https://doi.org/10.1063/1.5019779>
- [3] Unke, O.T., Chmiela, S., Gastegger, M. et al. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat Commun* **12**, 7273 (2021). doi: 10.1038/s41467-021-27504-0
- [4] Lookman, T., Balachandran, P. V., Xue, D. & Yuan, R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for construction of surrogate models. *npj Comput. Mater.* **5**, 21 (2019). <https://doi.org/10.1038/s41524-019-0153-8>

- [5] Liu, Haiyan et al. Curcumin potently blocks Kv1.4 potassium channels. *Biochem Biophys Res Commun* **344**,4: 1161–1165 (2006). doi:10.1016/j.bbrc.2006.04.020
 - [6] Aréchiga-Figueroa, Iván A et al. Modulation of Kv2.1 channels inactivation by curcumin.” *Pharmacological Reports*: **67**,6, 1273–1279 (2015). doi:10.1016/j.pharep.2015.05.019
 - [7] Fuchs, James R et al. Structure-activity relationship studies of curcumin analogues. *Bioorg Med Chem Lett* **19**,7, 2065–2069 (2009). doi:10.1016/j.bmcl.2009.01.104
 - [8] Seifi, M., Soltanmanesh, A., and Shafiee, A. Quantum coherence on selectivity and transport of ion channels. *Sci Rep* **12**, 9237 (2022). <https://doi.org/10.1038/s41598-022-13323-w>
 - [9] Engel, G. S. et al. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature* **446**, 782–786 (2007). doi: 10.1038/nature05678.
 - [10] Scholes, G. D., Fleming, G. R., Olaya-Castro, A. & van Grondelle, R. Lessons from nature about solar light harvesting. *Nat. Chem.* **3**, 763–774 (2011). <https://doi.org/10.1038/nchem.1145>
 - [11] Harel E, Engel GS. Quantum coherence spectroscopy reveals complex dynamics in bacterial light-harvesting complex 2 (LH2). *Proc Natl Acad Sci U S A.* 109(3):706-11 (2012). doi: 10.1073/pnas.1110312109.
 - [12] Krenn, M., Häse, F., Nigam, A., Friederich, P. & Aspuru-Guzik, A. Self-referencing embedded strings (SELFIES): A 100 % robust molecular string representation. *Mach. Learn. Sci. Technol.* **1**, 045024 (2020). doi:10.1088/2632-2153/aba947
 - [13] Jaeger, S., Fulle, S. & Turk, S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J. Chem. Inf. Model.* **58**, 27–35 (2018). doi: 10.1021/acs.jcim.7b00616.
 - [14] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *Proc. Int. Conf. Mach. Learn.* 1263–1272 (2017).
 - [15] Merz, Kenneth M Jr et al. Generative Models for Molecular Design. *Journal of chemical information and modeling* **60**,12, 5635–5636 (2020). doi:10.1021/acs.jcim.0c01388
-